

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2002, 05:48:11 ; Search time 117.02 Seconds
(without alignments)
14426.928 Million cell updates/sec

Title: US-09-840-743-5
Perfect score: 6873
Sequence: 1 gttctccggcattgactgc.....aaaaaaaaaactcgag 6873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	91.6	1.3	7218	1	US-08-232-463-14
C 2	73	1.1	7218	1	US-08-232-463-14
C 3	43.6	0.6	2207	4	US-08-956-322-3
C 4	43.6	0.6	2329	4	US-08-956-322-1
C 5	43	0.6	2140	2	US-08-314-601-1
C 6	43	0.6	2140	5	PCT-US95-13051-1
C 7	42.2	0.6	9636	1	US-08-323-170B-1
C 8	42.2	0.6	9636	4	US-08-954-441-1
C 9	41.4	0.6	19124	2	US-08-487-826B-13
C 10	40.6	0.6	289	4	US-09-007-005-17
C 11	40.6	0.6	289	4	US-09-244-796-17
C 12	40.6	0.6	1298	3	US-08-948-705-3
C 13	40.4	0.6	2447	2	US-05-014-969-14
C 14	40.2	0.6	1493	1	US-08-340-820-24
C 15	40.2	0.6	1493	1	US-08-593-535-24
C 16	40.2	0.6	1602	1	US-08-530-950-3
C 17	40.2	0.6	1602	3	US-08-888-429A-3
C 18	40.2	0.6	1602	4	US-09-149-879-3
C 19	40	0.6	6243	2	US-09-056-075-1
C 20	39.8	0.6	87350	3	US-08-781-891-79
C 21	39	0.6	277	4	US-09-007-005-3
C 22	39	0.6	277	4	US-09-244-796-3
C 23	39	0.6	289	4	US-09-007-005-17
C 24	39	0.6	289	4	US-09-244-796-17
C 25	39	0.6	306	4	US-09-461-697-203
C 26	39	0.6	696	4	US-09-461-697-193
C 27	39	0.6	699	4	US-09-461-697-191

28	39	0.6	717	4	US-09-461-697-189	Sequence 189, App
29	39	0.6	774	4	US-09-461-697-187	Sequence 187, App
30	39	0.6	819	4	US-09-461-697-185	Sequence 185, App
31	39	0.6	1669	4	US-09-461-697-184	Sequence 184, App
C 32	39	0.6	1797	4	US-09-088-140A-11	Sequence 11, Appl
C 33	39	0.6	3292	4	US-09-088-140A-12	Sequence 12, Appl
C 34	38.8	0.6	240	1	US-08-628-417-6	Sequence 6, Appl
C 35	38.8	0.6	1733	3	US-09-073-569-1	Sequence 1, Appl
C 36	38.8	0.6	6755	3	US-08-931-999-4	Sequence 4, Appl
C 37	38.6	0.6	282	4	US-09-461-697-205	Sequence 205, App
C 38	38.6	0.6	9636	1	US-08-323-170B-1	Sequence 1, Appl
C 39	38.6	0.6	9636	4	US-08-954-441-1	Sequence 1, Appl
C 40	38.4	0.6	1117	4	US-09-247-373B-33	Sequence 33, Appl
C 41	38.4	0.6	1956	4	US-08-559-896B-1	Sequence 1, Appl
C 42	38	0.6	277	4	US-09-007-005-3	Sequence 3, Appl
C 43	38	0.6	277	4	US-09-244-796-3	Sequence 3, Appl
C 44	38	0.6	1798	4	US-09-797-906-1	Sequence 1, Appl
C 45	38	0.6	1897	1	US-08-184-632-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pz29pt-Fls
US-08-232-463-14

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 4091 TTTTATATTACATGTAATCTACATTTTCATCAAAATGTTTTATTACTCTTGGATTA 4033

Db 4091 TTTTATATTACATGTATATTCTACATTTTCATCAATGTTTATTACTCTTGG


```

Qy      268 attgtcttttaccattcccccaaacatcggcttatcttcacgctcccttttta 328
        ||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     4091 TTTTATATTACATGTATATTCACATTTTTCATCAAAAGTGTTATTTACTCTGGATA 4033

RESULT          9
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.00ICPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match              0.6%; Score 41.4; DB 2; Length 19124;
Best Local Similarity    46.0%; Pred. No. 0.92;
Matches 176; Conservative 0; Mismatches 206; Indels 1; Gaps 1;

Qy      61 attcttggtttattacaataattcttcctttttctctctcccctttttatctcggaatt 120
        ||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     15871 ATTTATTATTTTTTTTTTATTTAATAAATTTTTTTTATTATTTATTTATTTAAAAT 15812

Qy      121 tgtctgcctaattttccagctgttacattttccgatcacgagaagaactcactgggt-tt 179
        - - - |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     15811 AAATTTTTTTTATTATGATATATATTTTTTTTTAACATTTTTTTTAAATTTTTTTTATT 15752

Qy      180 ttatgtaaatcaatacatgcttcctggtttcttgtatcataaaatcgcactattaacacctga 239
        |||| | |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     15751 TTATGATATATATTTTATTATTTTAATATATATTTTTTCTCTTTTTTTTGTGTTTATGATATA 15692

Qy      240 tttagattctgcgtaataaaaacctctgatttgctttttatcttcacctttccccaataaca 299
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     15691 TAATTTTTTTTTTTTAAATGTTTTTTTTTCTCTCTGTTGTTTTGTTTTTATTTTATTAATCA 15632
```

```

RESULT      8
US-08-954-441-1/c
; Sequence 1, Application US/08954441
; Patent No. 6316000
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,441
; FILING DATE: 20-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/323,170
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 015280-113110US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
; US-08-954-441-1

Query Match      0.6%; Score 42.2; DB 4; Length 9636;
Best Local Similarity 48.5%; Pred.No.0.36;
Matches 145; Conservative 0; Mismatches 153; Indels 1; Gaps 1;

QY 28 tcagaaagcgttagatcggtgagcgttttagctccattttctgttatttaccatatttc 87
Db 4330 TAAGAATACATAAAATGTAATAGATCCACTTTTCTTCATTATATAAAATTAACATTTC 4271
QY 88 cttttttctctcccttttttatctgggaatttggctctgctaaaaatttccagctgtac 147
Db 4270 TTTTGTTTATGTAACAACTTCTGGAATTAGGAATTAATTTATGTAATCTTCA-TTATTA 4212
QY 148 attttccgatcacgagagaatcacctgggttttttatgttaataacacatgttccctgtt 207
Db 4211 ATTTTAAATTTGGGACTCTGTTGGTTTACAAATCATATAATACATTGTATATTTT 4152
QY 208 tctgatacataaattcagcgtatttaaacacctgattttgtattctgcgtaataaaacctcg 267
Db 4151 TATAAATACCAAGACTTTTGGCTGGACAAAATATTTGAAATATAATTAAGAAATCTCTTA 4092

```

```

Qy 300 ttgcttactttattcgctcttctttaccgttttcacagctaaaaaatctctcgcgtattcaa 359
      || || || || || || || || || || || || || || || || || || || || ||
Db 15631 TTTTNTTTTATATAAAATTTTNTTTTAAATTTTNTTTTGATAATCTTTTTCATTTTTTAT 15572

Qy 360 tbggtttctcgtttgttgtagagaaaaatctatgcacaaaaatcattttatgcatttta 419
      || || || || || || || || || || || || || || || || || || || || ||
Db 15571 TCTATCAAAATTTATATATTTATATATATTTTATTTATTTTAAATAATTTTCTCCTTT 15512

Qy 420 tggtagcagattcttagttaatgt 442
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15511 TTTTNTTTTNTTTTNTTTTNTTTTATTTT 15489

RESULT 10
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 0.6%; Score 40.6; DB 4; Length 289;
Best Local Similarity 4.9%; Pred. No. 0.11;
Matches 12; Conservative 112; Mismatches 121; Indels 0; Gaps 0;

Qy 3445 acaggatgcataatctgtatctaggagacaagaagaagacacaaatgcaatgg 3504
      :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 9 rcrarururrcrururururrcrurururrcrurururrcrurururrcrururrcrnr 68

Qy 3505 tcttgacaaaggagatggtgcactgttccctatgagacgaagcgaaaaaccaagac 3564
      :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 69 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 128

Qy 3565 ccaagttgacattgacatgaacacactgcgatctgaactactgatggggaaggag 3624
      :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 129 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 188

Qy 3625 atgaaaaagaggggatgaagagaaggataaaaaagagaagtggtgggaagaagaaa 3684
      :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 189 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 248

Qy 3685 gaaga 3689
      :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 249 rnrgr 253

RESULT 11
US-09-244-796-17

```

```

; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 0.6%; Score 40.6; DB 4; Length 289;
Best Local Similarity 4.9%; Pred. No. 0.11;
Matches 12; Conservative 112; Mismatches 121; Indels 0; Gaps

Qy 3445 acaggatgcacaaatctgtatcttagggagacaaagaagagacaaagcaaatgcaatgg 3504
   :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 9 rcratarururcrurururururururururururururururururururururur 68

Qy 3505 tctgtacaaggagatggcactgttcctatgagagcaagaagcgaaacccaagac 3564
   :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 69 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 128

Qy 3565 ccaaagttgcattgcagatgaacacgcgcatactggaacttactgtggggaaggag 3624
   :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 129 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 188

Qy 3625 atgaaaagaagggatgaagagaagagataaaaaagaagaagtggtgtgggaagagaa 3684
   :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |:
Db 189 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 248

Qy 3685 gaaga 3689
   :|:
Db 249 rurgr 253

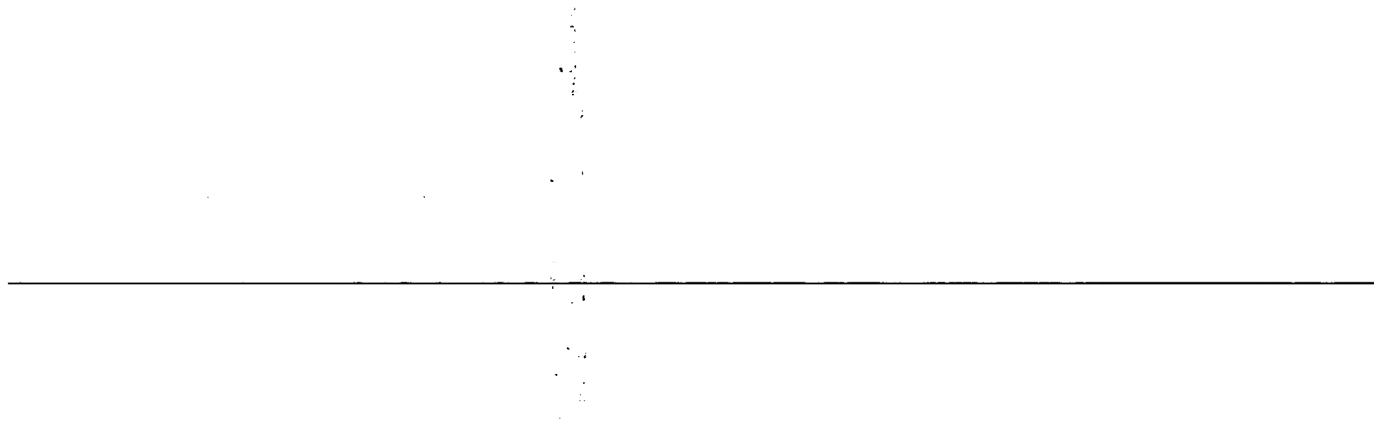
RESULT 12
US-08-948-705-3
; Sequence 3, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Chen, Yao-Tsung
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSIS
; FILE OF INVENTION: TREATING COLON CANCER
; FILE REFERENCE: LUD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/08/948,705A
; CURRENT FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 8

```

RESULT 11
US-09-244-

QY	3352	aaggaaaagggaagtccatcacagattcaggaaaaagcaagaggtccatcaggagaac	3411
Db	903	aagaaagaagaagagaggaaggaagaaagaaaggaaggaaggaagga	962
QY	3412	tctgtgtcaggattctattgcggaataaattcacagatgcatactgtatctcaggag	3471
Db	963	gaaagaagaaggagaagaaaaggaaaaagagaagaaagaaactaagaagaagaggag	1022
QY	3472	acaagaagaagagacaaagagcaaaatgcaatgtgtctgtacaaaggagatggtgcacttg	3531
Db	1023	aggaagataaagaagaaagagaagaaagaaagatnaaagaagaaagaaagaaagaa	1079
QY	3532	tcccttatgagcagcaagcgaaaaaccacagaccgaattgacatgacgatgaacaa	3591
Db	1080	aaggaaaagagaggagaactnagaagaagaagagaggagaagaaagaaagaataaag	1139
QY	3592	ctcgcatacggaaacttactgatcggggaaagagatgaaaaagaaggagatgaagagaagg	3651
Db	1140	gaacnagaagaagagagaagaaagaataaagaaagggaagaaagaaagaaagaaagaa	1199
QY	3652	ataaaagaagaagagtggtgtgggaagagaagaaagaa	3689

RESULT 14
US-08-340-820-24/c
; Sequence 24, Application US/08340820
; Patent No. 5512460



THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2002, 05:47:46 ; Search time 8056.51 Seconds
(without alignments)
17852.406 Million cell updates/sec

Title: US-09-840-743-5
Perfect score: 6873
Sequence: 1 gttctccgcatgactgcg.....aaaaaaaaaaactcgag 6873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_em:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description

1	4790.2	69.7	93045	8	ATT32M21	AL162875 Arabidops
2	3895.2	56.7	83485	8	ATTIE3	AL162972 Arabidops
3	290	4.2	6055	8	ATSULFRED	Y10157 A.thaliana
4	182.2	2.7	361	8	AB051105	AB051105 Euphorbia
5	117.8	1.7	101365	8	AC006919	AC006919 Arabidops
6	100.4	1.5	97366	8	AP002745	AP002745 Oryza sat
7	100.4	1.5	134967	8	AP002861	AP002861 Oryza sat
8	91.6	1.3	7218	6	I66494	I66494 Sequence 14
9	79.4	1.2	160	8	AF462201	AF462201 Narcissus
10	73	1.1	7218	6	I66494	I66494 Sequence 14
11	66.2	1.0	863	11	CNS06EVO	AL395628 T7 end of
12	66	1.0	152988	2	AC073669	AC073669 Mus muscu
13	65	0.9	5750	6	AX251465	AX251465 Sequence
14	65	0.9	5750	6	AX346910	AX346910 Sequence
15	63.2	0.9	685	11	CNS06EJ4	AL395174 T3 end of
16	63.2	0.9	96232	8	ATAC010927	AC010927 Arabidops
17	62	0.9	49885	6	AX059537	AX059537 Sequence
18	62	0.9	53778	8	TLJ1	AF128393 Arabidops
19	62	0.9	198563	8	ATCHRIV14	AL161502 Arabidops
20	61.4	0.9	63365	2	AC087437	AC087437 Homo sapi
21	61.4	0.9	143585	2	AC013349	AC013349 Homo sapi
22	60.2	0.9	16033	6	AX346306	AX346306 Sequence
23	60.2	0.9	81120	2	AC022851	AC022851 Homo sapi
24	60.2	0.9	127709	2	CNS05TC1	AL352983 Homo sapi
25	60	0.9	222068	2	AC087188	AC087188 Homo sapi
26	59.4	0.9	146936	2	AC106157	AC106157 Rattus no
27	58.8	0.9	6145	6	AX345874	AX345874 Sequence
28	58.8	0.9	42612	2	AC091325	AC091325 Mus muscu
29	58.8	0.9	83440	2	AC024285	AC024285 Homo sapi
30	58.8	0.9	117193	2	AC099851	AC099851 Homo sapi
31	58.8	0.9	349980	6	AX344566	AX344566 Sequence
32	58.4	0.8	14006	6	AX346860	AX346860 Sequence
33	58.2	0.8	64612	2	AC079333	AC079333 Homo sapi
34	58.2	0.8	130540	2	AC079417	AC079417 Mus muscu
35	58	0.8	6419	6	AX345169	AX345169 Sequence
36	58	0.8	169743	2	AC086448	AC086448 Rattus no
37	57.6	0.8	56693	2	AC084077	AC084077 Homo sapi
38	57.4	0.8	131271	2	AC015927	AC015927 Homo sapi
39	57	0.8	6668	6	AX346598	AX346598 Sequence
40	57	0.8	53932	2	AC023371	AC023371 Homo sapi
41	57	0.8	78578	2	PFMAL13P9	AL096783 Plasmodiu
42	57	0.8	130808	2	AC026941	AC026941 Homo sapi
43	56.8	0.8	69284	2	AC023185	AC023185 Homo sapi
44	56.8	0.8	156550	2	AC015830	AC015830 Homo sapi
45	56.6	0.8	73778	6	AX344562	AX344562 Sequence

ALIGNMENTS

RESULT	1	LOCUS	ATT32M21	93045 bp	DNA	linear	PLN 31-MAR-2000
DEFINITION	Arabidopsis thaliana DNA chromosome 5, BAC clone T32M21 (ESSA project).						
ACCESSION	AL162875						
VERSION	AL162875.1						
KEYWORDS	GI:7406444						
SOURCE	thale cress.						
ORGANISM	Arabidopsis thaliana						

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 93045)
Bavan, M., Terryn, N., Ardiles, W., Buysshaert, C., Dasseville, R., De Clerck, R., De Keyser, A., Neyt, P., Rouze, P., Van Den Daele, H., Villarroel, R., Gielen, J., Van Montagu, M., Bancroft, I., Mewes, H.W., Rudd, S., Lemcke, K. and Mayer, K.F.X.

Unpublished
2 (bases 1 to 93045)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (31-MAR-2000) MIPS, at the Max-Planck-Institut fuer

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk

COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES

source

1..93045 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 1..1000
 /note="this clone overlaps with BAC T19M18"
 676..3112
 /gene="T32M21_10"
 join(676..841,958..1235,1354..1545,1884..2447,2525..2642,2727..3112)

/gene="T32M21_10"
 /note="strong similarity to NAC2, Arabidopsis thaliana, EMBL:AF201456"

/codon_start=1

/product="putative protein"

/protein_id="CAB85547.1"

/db_xref="GI:7406445"

/translation="MGRGVTSLAPGRFHTDEELVRYLLKRVCKNPKPKFAISVTT
 DTKSEPWDLPKSLKSLRDLEWFFSMLDKYSNGSKTNRATEKGWTKTQKGDREIR
 NGRSVGMKTLVYHKGRAPERTNNVMEHYRLSDLDLKAGVPQEAAYVLCRIFOKS
 GTGPKNGEYQAPYLEEWEEDGMTYVPAQDAFSEGLANDVDVVDIDIDEKENLV
 VYDAVPLIPNYCHGESNNVESGNYSDSGNYIQGNVVDGGYFEOPIETFEEDRKP
 IIREGSTQPCSLPEQIGCGVODENVNLESNNNVFVADTCYSDIPIDHNYLPDEP
 FMPDNNPLPLNDGLYLETNDLSAQQDDNFEDYLSFFDEGLTFDDSLMLGPEDFLP
 NQALQDPAPKELEKEVAGGKAVEKESEGSSKQDDFDKDFDAPKYPFLKKTIS
 HMLGAIPTPSFASQFQTKDAMRLHAAQSGSVHVYTAGMMRIISNMTLAADSGMGSYD
 R"
 KGNLNLVLSFGVYQQDDAMTASGSKTGITATRAMLVFMCVLWVLLLSVSFKIVTWSA

676..841
 /gene="T32M21_10"

/number=1

842..957
 /gene="T32M21_10"

/number=1

958..1235
 /gene="T32M21_10"

/number=2

1236..1353
 /gene="T32M21_10"

/number=2

1354..1545
 /gene="T32M21_10"

/number=3

1546..1883
 /gene="T32M21_10"

/number=3

1884..2447
 /gene="T32M21_10"

/number=4

2448..2524
 /gene="T32M21_10"

/number=4

2525..2642
 /gene="T32M21_10"

/number=5

2643..2726
 /gene="T32M21_10"

/number=5

2727..3112
 /gene="T32M21_10"

/number=6

complement(3563..3721)
 /gene="T32M21_20"

gene

/number=1

complement(join(3563..3721,3811..3870,3949..4269,
 4349..4432,4513..4601,4709..4823,4900..5026,5142..5281,
 5368..5429,5507..5573,5652..5783,5870..5947,6041..6151))
 /gene="T32M21_20"

gene

3563..6151

/gene="T32M21_20"
 complement(join(3563..3721,3811..3870,3949..4269,
 4349..4432,4513..4601,4709..4823,4900..5026,5142..5281,
 5368..5429,5507..5573,5652..5783,5870..5947,6041..6151))
 /gene="T32M21_20"

/note="similarity to rugB protein, Dictyostelium

discoideum, PIR:S68824"

/codon_start=1

/product="putative protein"

/protein_id="CAB85548.1"

/db_xref="GI:7406446"

/translation="MEEIRKEIDVGDWHSNLAHDEWTPLPVSGRSASARYKHAADVVD
 EKLYIVGSRNGRYLSDVOVDFDLRLTWSLKLKTESSADNIOEDDGSLSREAFPAI
 SDHRMLKGNKLLIGGHSKSSDNMLVRFIDLETHSCGVIDVFGNVPASRGGHSITL
 VGSRLVFGGEDKNRLLNDLHLHLETMVDVETKQTRPVPFRDHTAATHSDRIILL
 IFGCSHSIFYSDLHILDLOTMEWSQPHVQGVVTPRAGHAGITIDENWYIVGGDINS
 TGCLETIVLNMKSLVWSTHVEARHPLASEGLSVCSASVFGENILVAFGNYKGYNN
 DIFVRLKPGESSHPKIFKSPAAAAAASVTAAYATAKSDKSDYPPANPTLNGNGNS
 LPERDIRNRDITKEEKRALESSIAETOVENAKLRKIDENVSSHTLSQELQSVEGO
 LISERSRCFKLEAQIAELQKALESGQISIAEVEMLRQRORSASDEEDGTVQRGSAGV
 WGLFGR"

complement(3722..3810)

/gene="T32M21_20"

/number=1

complement(3811..3870)

/gene="T32M21_20"

/number=2

complement(3871..3948)

/gene="T32M21_20"

/number=2

complement(3949..4269)

/gene="T32M21_20"

/number=3

complement(4270..4348)

/gene="T32M21_20"

/number=3

complement(4349..4432)

/gene="T32M21_20"

/number=4

complement(4433..4512)

/gene="T32M21_20"

/number=4

complement(4513..4601)

/gene="T32M21_20"

/number=5

complement(4602..4708)

/gene="T32M21_20"

/number=5

complement(4709..4823)

/gene="T32M21_20"

/number=6

complement(4824..4899)

/gene="T32M21_20"

/number=6

complement(4900..5026)

/gene="T32M21_20"

/number=7

complement(5027..5141)

/gene="T32M21_20"

/number=7

complement(5142..5281)

/gene="T32M21_20"

/number=8

complement(5282..5367)

/gene="T32M21_20"

/number=8

complement(5368..5429)

		/gene="T32M21_20"	
		/number=9	
intron		complement(5430..5506)	
		/gene="T32M21_20"	
		/number=9	
exon		complement(5507..5573)	
		/gene="T32M21_20"	
		/number=10	
intron		complement(5574..5651)	
		/gene="T32M21_20"	
		/number=10	
exon		complement(5652..5783)	
		/gene="T32M21_20"	
		/number=11	
intron		complement(5784..5869)	
		/gene="T32M21_20"	
		/number=11	
exon		complement(5870..5947)	
		/gene="T32M21_20"	
		/number=12	
intron		complement(5948..6040)	
		/gene="T32M21_20"	
		/number=12	
exon		complement(6041..6151)	
		/gene="T32M21_20"	
		/number=13	
gene		complement(join(7298..7394,7484..7545,7615..7710, 8136..8197,8761..8954,9380..9486,9592..9830,10135..10219))	
		/gene="T32M21_30"	
		complement(7298..7394)	
exon		/gene="T32M21_30"	
		/number=1	
gene		7298..10219	
		Query Match 69.7%; Score 4790.2; DB 8; Length 93045;	
		Best Local Similarity 92.7%; Pred. No. 0;	
		Matches 5229; Conservative 0; Mismatches 13; Indels 401; Gaps 3;	
Qy	1	gtctccggcattgactcgctcgagaaacagccttagatcggtgagcttttagctcc	60
Db	65889	GTTCPCGGCATTGACTCGCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTAGCTCC	65948
Qy	61	atttctcttattacatatttcccttttttctctctcccttttttctctggaatt	120
Db	65949	ATTTTCTGTTTATTTACATATTTATTCCTTTTCTCTCTCCCTTTTATCTGGAATT	66008
Qy	121	tgtctctaaatttccagctgttacatttccgatccagagaagaactcctgggtttt	180
Db	66009	TGTTCTGCTAAATTTTCAGCTGTTTACATTTTCCGATCACCAGAGAATCCTACTGGGTTTT	66068
Qy	181	tatgtaatacatatgttctctgttttctgtatcataaattcagctattacacctgat	240
Db	66069	TATGTTAATCAATACATGTTCTGTTTCTGATCATATAATCTCAGCTATTTAACACCTGAT	66128
Qy	241	tttgattctcgtaataaaaacctctgtattgcttttattcttctactttcccccataaacat	300
Db	66129	TTTGTATCTCGGTAATAAAAACCTCTGATTTGCTTTTATCTTCACTTTCCTCCCAATAACAT	66188
Qy	301	tgccttaatttctgcgtcttcttttaacogtttcaagctaaaaaaattcttgcctattcaat	360
Db	66189	TGCTTACTTTTATTCGCTCTCTTTTACCGTTTCCAGCTTAAAAAATTTCTTCGCTATTCAAT	66248
Qy	361	gtgttctctgttttgttgatgagaaaaatactgacaaaaaatcatttattgcattttat	420
Db	66249	GTGTTTCTCGTTTGTGATGAGAAAAATATCTGNCAAAAAATCATTTATTGCAATTTTAT	66308
Qy	421	ggtgcagattcttagttaattgtcgcttctcttaaccaagtcagattaaaaaggagtgctc	480
Db	66309	GGTGCAGATTCTTACTTAATGTCGCTTCTCTAACCAAGTCAGATTAAAAAGGAGTGTC	66368
Qy	481	gtccatgttctgtttgttggagagagtttccgagaggttaggtgagtgattat	540
Db	66369	GTCCATGTTGCTTTGTTTGGTGGTGGAGAGAGTTTTCGGAGAGTTAGGTGAGTGTAT	66428

Qy	541	ttggggtgagtagtgataaggttttgaagggggagtgattcatcaagtgtgttatgaatt	600
Db	66429	TTGGGTTGAGTACTGATAAGGTTTGAAGGGGAGTGATTATCAAGTGTCTTATGAATT	66488
Qy	601	cgagggtgatccgggggagatagatatatttccagttcccttggagaaatacaaaactcaacaag	660
Db	66489	CGAGGGTGTATCCGGGGGATAGATATTTTCAGTTCCTTTGGGAATCAAACTCAACAAG	66548
Qy	661	agttcatgggttcttgattccatttacaccacaaaaccttagatcaagtctgtggttag	720
Db	66549	AGTTTATGGGTTCTTGGATTCCATTACACCCAAAAACCTAGATCAAGTCTGTATGGTAG	66608
Qy	721	atgagagagtataaacacagatctaaatgggtttccagggtgggtgaattttgtagacaggg	780
Db	66609	ATGAGAGAGTGATAAACACAGGATCTAAATGGTTCAGGTGGTGAATTTCTAGACAGGG	66668
Qy	781	gattctgcaacactgggtgtgagatcataatgggggttttttgatcatcagggtcagggcg	840
Db	66669	GATTCTGCAACACTGGTGTGGATCATATATGGGGTTTTTGATCATCGTCTCATCAGGCG	66728
Qy	841	ttaccaacttaagtatgatgatcaatagcttagcggatcacatgcacaagcttggagta	900
Db	66729	TTACCAACTTAAGTATGATGATCAATAGCTTAGCGGGATCACATGCACAAGCTTGGAGTA	66788
Qy	901	atagttagagagatcttcttggcaggagtgagtgactctctcttcccttttagcaccagttatca	960
Db	66789	ATAGTGAGAGAGATCTTTTGGCGAGGAGTGAGGTGACTTCTCTCTTAGCACCAGTTATCA	66848
Qy	961	gaacacacacccggtaattgtagagccggtcaatggaaatttttacttcagatcgggtatgg	1020
Db	66849	GAACACACCCGGTAATGTAGAGCCGGTCAATGGGAAATTTTACTTTCAGATCGGTATGG	66908
Qy	1021	taaatgtcccttccaccagagtgccactctcaagctcggctataatgaagtgttaattgg	1080
Db	66909	TAAATGGTCCCTTCCACCAGAGTGGACTTCTCAAGCTGGCTATAATGAGTTGAATTTGG	66968
Qy	1081	atgacttgttaactcctgatcagatgcccttctccttcacaaagcttgcgtgagtggtggg	1140
Db	66969	ATGACTTGTTCATCCITGATCAGATGCCCTTCTCTTCCACAAGCTTGTGAGTGGTGGG	67028
Qy	1141	atagcttattcaagggttcgaatgtgagtgatcaaatctatttctcagtttttttttttc	1200
Db	67029	ATAGCTATTTCGAAGGTTGTCATATGTGAGTATCAAAATCTATTTTTCAGTTTTTTTTTC	67088
Qy	1201	ccttctccgttcttcagtlacttagtagtagaagacatgaatagaatatcttaagaaagt	1260
Db	67089	CCTTCTTCCCGTCTTCGAGTACTTAGAGTAGAATGAATAGAAATATCTTAGAAAAGT	67148
Qy	1261	catggttttgacagatgagacctccagctgtaacaaagcctctttacaaatttgattcac	1320
Db	67149	CATGGTTTGAACAGATGGACCTCCAGCGTCTAACAAAGCTCTTTACAAATTTGAATTCAC	67208
Qy	1321	caattagaagaagacagttgggttcagtcgtgaaagttcgtttcaatatgtacogtcaa	1380
Db	67209	CAATTAGAAGAAGACAGCAGTTGGGTGAGTCTGTGAAAGTTCGTTTCAATATATGACCGTCAA	67268
Qy	1381	cgccagctctttcagaacaggtgaaaagactggattccttgaacagataagttacaacta	1440
Db	67269	CGCCAGCTCTTTCAGAACAGGTGAAAGACTGGATTCTTGAACAGATAGTTTCAACTA	67328
Qy	1441	ctggacatgaaatcccagagccgaaatctgacaaaagtatgcagagcatttatggactcgt	1500
Db	67329	CTGACATGAATCCAGAGCCGAAATCTGCACAAAGTATGCAGAGCATATTGGACTCGT	67388
Qy	1501	ctgctgttaatgcagcgaagctactgaacaaaatgatggcagcagacaagatgtctcgg	1560
Db	67389	CTGCTGTTAATGCGACGGAAGCTACTTGAACAAATATGATGCGCAGCAGACAAGATGTTCTGG	67448
Qy	1561	agttcgaacttaacaaaactcctcagcagaaaacccctcaaaaggaaaaggaggttcacgc	1620
Db	67449	AGTTTCGACCTTAACAAAACCTCTCAGCAGAAACCCCTCCAAAGGAAAAGGAAGTTCATGC	67508

Db 69669 AAGAAGAGCTTCCGAGGAAGGGCTGATTCCTTCATCGCTCGCATGCACCTGGTACAAAGG 69728
Qy 3740 ----- 3739
Db 69729 TGAAGATCCACTTCTCTCTCAACTCCATTTTATTACACACAAATTAGTAGAATACTCAA 69788
Qy 3740 ----- 3739
Db 69789 AAATGATGTTTTGTTGCAAAATTTTAAATTCATCTAGTTAACCATGTCAAATATATTC 69848
Qy 3740 ----- 3739
Db 69849 ATAATGCATCTTGTGAACACAGGTGTCATTTATGTGTACAGCTGAATGGTTTTATGTGC 69908
Qy 3740 -----aggagataga 3749
Db 69909 CTATTATTCTTTTACTGCTATAGATGACCAATTTGAACCTTAAACGTTTACAGGAGATAGA 69968
Qy 3750 cgtttttcccatggaaggagatcgggttgattcggttcattcggttcatttccttcacacag 3809
Db 69969 CGTTTTGCGCCATGGAAGGGATCGGTGATTTCGCTCATTTGGAGTTTTCCTTTACACAG 70028
Qy 3810 aatgtctcggtaccccttctcc----- 3830
Db 70029 ANTGTCTCGGATCACCTTTCAGGTATATAGATTGCTTAAATAAATTCAGTTGCCAAAACA 70088
Qy 3831 -----agctctgcgttcattgctctctagctgctcg 3859
Db 70089 TAGAAATTAACCCATGTTGGTTTACAAATGCAGCTCTGCGTTTCATGCTCTAGCTGCTCG 70148
Qy 3860 attccctccaaaattagcagcagccgagagaatgaaagggaattgttagaagcgttagttgt 3919
Db 70149 ATTCCCTCCAAAATTAAGCAGCAGCCGAGAGATGAAAGGAATGTTAGAAGCGTAGTTGT 70208
Qy 3920 tgaagatccagaagatgctcttgaaccttaaatgaaattccttcgtgaggaagaaagt 3979
Db 70209 TGAAGATCCAGAGATGCATCTGAACTTAATGAAATTCCTTCGTGGCAGGAAAGGT 70268
Qy 3980 tcaacatccatctgacatggaagttctctgggttgatagtgatcacaagagcgactaaag 4039
Db 70269 TCAACATCCATCTGACATGGAAGTTTCTGGGTTGATAGTGATCANAAGAGCAGCTAAG 70328
Qy 4040 ggaactgttcaaacctcgtgaattgaaagatttaattcttttagagaagagatattcaaaattt 4099
Db 70329 GGACTGTTTCAAACCTCGGAATTTGAAGATTTAATTTCTTAGAGAAGAGTATTCAAATTT 70388
Qy 4100 agaagagagaagtattatcatcacagaatctcttgatccgcgcgatattcagtcgtgtg 4159
Db 70389 AGAAGAGGAAGTATTATCATCACAAGATCTTTTGTCCGCGGATATTTCAGTCGTGTGG 70448
Qy 4160 gagagttggtacctgttctatgttccaaatcagacgcagagtttctctacaacccaggtgtga 4219
Db 70449 GAGAGTTGGATCCCTGTTTCATGTTCCAAATCAGACGAGAGTTTCTTACAACCCAGGTGTA 70508
Qy 4220 aacaaaactcagtggaacatcacaaatcagtgcaaacctggaagcgggagccaaactgtctga 4279
Db 70509 AACAAAACCTGTCTAGTGAACATCACAAATCAGTGCACAACTGGGAGCCAAACTTGTCTGA 70568
Qy 4280 tgaattgtctcaagggaatgagaccgcacatctatatgaaggatctggtgatgttca 4339
Db 70569 TGAATTTGCTTTCAGGGAATGAGACCCGCATCTATATGAAGGATCTGTTGATGTTCA 70628
Qy 4340 gaaacagaactcaaatgctgcgtcagaagaacactgatcttgaaaaaacaatgaattg 4399
Db 70629 GAAACAGAACTACAATGTCTGTCTAGAGAAACCTTGATCTTGAAAAACAATGAATTG 70688
Qy 4400 gaaagactctgtctgtttgttcagcgcgaagaatgtactactaattgtggcaacaactcttc 4459
Db 70689 GAAAGACTCTGCTGTTTGTGTGTCAGCAAGAAATGATACTAATTTGGCAACAACCTCTTC 70748
Qy 4460 cagcagctatgagcagtgctgcgactcagacagccacatgtactagacatagaggtatttg 4519
Db 70749 CAGCAGCTATGAGCAGTGTGCGACTCGACAGCCACATGTACTAGACATAGAGGATTTTGG 70808

Qy 4520 aatgaagggtgaagcccttggttattcttggatgccatctcaccagaggttgacagagt 4579
Db 70809 AATGCAAGGTGAAGGCCCTTGTTATTCTTGGATGTCATCTCACCAGAGTGTACAGAGT 70868
Qy 4580 aagaacacaaaaatgtaccacgcaggttttttcagacaagggtggaagtgttccaagagaatt 4639
Db 70869 AAGAACAACAAAATGTACACGCAAGGTTTTTCAGACAAGGTGGAAGTGTTCACAGAGAATT 70928
Qy 4640 cacaggctcagatcatcatcaacgcctcatgaattaccaggaatggattgctcgggttc 4699
Db 70929 CACAGGTGAGATCATATCATCAACGCTCATGAATTACCAAGATGGATTGTCCGGTTC 70988
Qy 4700 ctcaagcgcgttccaagaacacaccaggagcataccacaataatcaacaagaatgatgaa 4759
Db 70989 CTCGAAGCGCGTCCCAAGAACACCAGGACGATACCCAACATAATCAACAAGATGAGATGAA 71048
Qy 4760 taaagcatcccatattacaacaaaaacatttttggatctcctcaactcctctgaagaatgcct 4819
Db 71049 TAAAGCATATCCCATTTTACAAAACATTTTGGATCTGCTCAACTCCTCTGAAGAATGCCT 71108
Qy 4820 tacaagacagtcacagtcaccaacacagacatcacggtggtgctaccgagagatagaac 4879
Db 71109 TACAAGACAGTCCAGTACCAACAGACATCACGGATGGCTGCTCTACCGAGAGATAGAAC 71168
Qy 4880 tgcgtgaagacgtggttgcctgcctcagtaacaattcaagcttaccagacaatatattggtcga 4939
Db 71169 TGTGTAAGACGTGTTGATCCGCTCAGTAAACAATTCAAGCTTACAGAACATATTTGGTCGA 71228
Qy 4940 atcaaatccagcaataaagaagcagcaggttggaatacaagaagagacaaatgcactat 4999
Db 71229 ATCAAAATCCAGCAATTAAGAGCAGACGCGAGTTGAAATACAAAGGAGACAAATGCCACTAT 71288
Qy 5000 ttacgagagatgaaggagcgtctgctgaggaagaaagcctacaagccagtggtgtag 5059
Db 71289 TTTACGAGAGATGAAGGAGCGCTTGTGATGGGAAAAGCCTACAAGCCAGTGGATAG 71348
Qy 5060 tctcagaaaaagatgtgaggggaatgaaggagagagagacgaacgaacaaaaacaatatgga 5119
Db 71349 TCTCAGAAAAGATGTGGAGGGGAATGAAGGGAGACAGAGCAACAAAACAATATGGA 71408
Qy 5120 ttccatagactataagaacaataagacgtgctagtagtatcagcgagatttctgagggtatcaa 5179
Db 71409 TTCCATAGACTATGAACCAATAAGACGTGTAGTATCAGCGAGATTTCTGAGGCTATCAA 71468
Qy 5180 ggaagagagatgaataacatgttggcgtacgaattcaagatttcttagaacggatag 5239
Db 71469 GGAAGAGGGATGAATACATGTTGGCCGTACGAATTAAGGTAATCTACTAATTTTCAGT 71528
Qy 5240 taa 5242
Db 71529 TGA 71531

RESULT 2

ATTIE3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

Arabidopsis thaliana DNA linear
Arabidopsis thaliana DNA chromosome 5, BAC clone T1E3 (ESSA project).
AL162972.1 GI:7413526thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis
1 (bases 1 to 83485)
Bevan, M., Terry, N., Ardiles, W., Buyssehaert, C., Dasseville, R., De
Clerck, R., De Keyser, A., Neyt, P., Rouze, P., Van Daele, H.,
Villarroel, R., Gleien, J., Van Montagu, M., Bancroft, I., Mewes, H.W.,
Rudd, S., Lemcke, K. and Mayer, K.F.X.
Unpublished

REFERENCE	2 (bases 1 to 83485)	gene	complement(join(27911..28137,28504..28621,28713..30023,30153..30797))
AUTHORS	EU Arabidopsis sequencing,project..		/gene="TIE3_30"
TITLE	Direct Submission	CDS	complement(join(27911..28137,28504..28621,28713..30023,30153..30797))
JOURNAL	Submitted (03-APR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk		/gene="TIE3_30"
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.		/note="similarity to BRL protein, Homo sapiens, EMBL:AF005067"
FEATURES	Location/Qualifiers		/codon_start=1
source	1. 83485		/product="putative protein"
	/organism="Arabidopsis thaliana"		/protein_id="CAB86009.1"
	/variety="Columbia"		/db_xref="GI:7413529"
	/db_xref="taxon:3702"		/translation="MPSVGMRRTTRVGVVKAADGARVLRSGRRIPWNPGEKPVRAH DVIDRDCDVLKQNKNGKNGKSSQPCSPKQVSEKEDKVDVDFVTKRKRVRN EGVGDEKTVDMKFGIVYSKRKRKLCPESSSDRSEELRSILKFYRRRKLRSQVSVLT LVWDSCDCWFLTVFGLAMRYIRREELRLSSLASFLSQPINQVADHGVRLYRSP LSRGVCKFFGAMSCPLPLFSADFAVIRPMFMHFTLFRVLPDRSFFVEKSLYLNN PIESDSSELALPECTPRNGVVYGLHPSVRASKLTGNAQRNGLSGHSFQKRKRS LRRRARNLSHNAKLNNGTIPVFDISGRKNRTAAVSSKKLRSVLSNPSVNGIS IPMTKTKEELDSICCSANILMIHSDRCTREGEFSVMLEASSKSEWELVIKDCGAIYS HMAQRTMRPFSSNRITHATVWGGDNWKLFCDDQDLGFKDIYKCYERNLLEQSVK VTIPGVREVCGVYAEYIDNPFSPSPVSVISVNEDEVSRAMARSTALYDMDSDEW LERONKMLNEEDDQVLOQOREAFELMIDGFEKYPHPSADLLDKAATIGSISYLG ROEYVEAVHDYWLKRRKQKRAPLLRIFQGHQVKKTLQLLSKPVPFRKRFRKQSQSLHG KAKQTSPPMWAVKAAPEEDEDLIRMEEAQVLADTKTMTAKRRRAQILAEADLAV YKAMRALRIAEAIKEAESREVDTVV"
	/chromosome="5"	exon	complement(27911..28137)
misc_feature	1. .26282		/gene="TIE3_30"
	/note="overlap to BAC T32M21, EMBL:AL162875; please refer to this entry for analysis and annotation"		/number=1
exon	24316..24381		27911..30797
	/gene="TIE3_10"	gene	/gene="TIE3_30"
gene	24316..24869	intron	complement(28138..28503)
	/gene="TIE3_10"		/gene="TIE3_30"
CDS	join(24316..24381,24630..24752,24849..24869)	exon	/number=1
	/note="similarity to predicted protein, Arabidopsi thaliana"		complement(28504..28621)
	/codon_start=1		/gene="TIE3_30"
	/product="putative protein"	intron	/number=2
	/protein_id="CAB86007.1"		complement(28622..28712)
	/db_xref="GI:7413527"		/gene="TIE3_30"
	/translation="MAGKYWEKTPCNWFKCNVGGAWYKMWLVLETFGANRANLIAS VTGDLRLQSYVANGPPSCLSVVEFV"	exon	/number=2
intron	24382..24629		complement(28713..30023)
	/gene="TIE3_10"		/gene="TIE3_30"
exon	24630..24752		/number=3
	/gene="TIE3_10"	intron	complement(30024..30152)
	/number=1		/gene="TIE3_30"
intron	24753..24848	exon	complement(30153..30797)
	/gene="TIE3_10"		/gene="TIE3_30"
	/number=2	CDS	complement(join(36227..36865,36944..37374,37440..37631,37768..38448,38589..38763))
exon	24849..24869		/gene="TIE3_40"
	/gene="TIE3_10"		/note="similarity to prtected proteins, Arabidopsis thaliana
gene	25962..27500		Contains Prokaryotic membrane lipoprotein lipid attachment site AA660-670"
	/gene="TIE3_20"		/codon_start=1
CDS	25962..27500		/product="putative protein"
	/gene="TIE3_20"		/protein_id="CAB86010.1"
	/note="strong similarity to cytochrome P450 77A3p, Glycine max., PIR:T05948		/db_xref="GI:7413530"
	Contains Cytochrome P450 cysteine heme-iron ligand signature AA449-458"		/translation="MAISSNYLHATLMATLRSPKFAANTKVPHADLGSVSPTKKNP EEVPDNPVNIQLRGKTNVYVEYIQLSOGISQGRVVEVTKDFLNDHPDAVDWNLVE TPLLKACAGKPEIKELLMRMPPEMLPKMSQNSVHTATLVYAVSGNWEIAEALVA RQSAMDKLIRIKSEARDNDGLIWLKNARFAACILLCEMCEIEMDEETIEKMDLIL KTVLMVDRIDDLYLILAPFFSKERKRALEVRQGVYGVIERRAIINQSGDK TASSFSYLDLFLDKLEKRTTPSNEELVLCSEFLNGDGTGTAIEWGIAQLIANP EQSRDYLDIKSTGVDRVDEKVDKWFLOAFKELLRKHPPTVFSITHAVMETTT LAGYDIPAGVNVVYLPGISDPRINWNPKFDPDRMLCKEDADITGISGVMIPFG VGRICPGLAMATIHVHLARMVQEFWCAHPGSEIDFAGKLEFTVVMKNPLRAMV KPRI"
	/codon_start=1		/product="cytochrom P450-like protein"
	/protein_id="CAB86008.1"		/protein_id="CAB86008.1"
	/db_xref="GI:7413528"		/db_xref="GI:7413528"
	/translation="MEPLISFSPTSLSLDTFFAIILISGFFVFIITRWNSNKKRLNLP PGWVGLNFARSGKPPFEYAEADLKKTYGIFTLRMGTRTMIILSDATLVHEALI QRCALFASPAENPTRTFSCNKFTVNAKYGPVWRSRLRNVMQNLMSRLKFEKGL RQSAMDKLIRIKSEARDNDGLIWLKNARFAACILLCEMCEIEMDEETIEKMDLIL KTVLMVDRIDDLYLILAPFFSKERKRALEVRQGVYGVIERRAIINQSGDK TASSFSYLDLFLDKLEKRTTPSNEELVLCSEFLNGDGTGTAIEWGIAQLIANP EQSRDYLDIKSTGVDRVDEKVDKWFLOAFKELLRKHPPTVFSITHAVMETTT LAGYDIPAGVNVVYLPGISDPRINWNPKFDPDRMLCKEDADITGISGVMIPFG VGRICPGLAMATIHVHLARMVQEFWCAHPGSEIDFAGKLEFTVVMKNPLRAMV KPRI"		
exon	25962..27500		25962..27500
	/gene="TIE3_20"		/gene="TIE3_20"
	/number=1		/number=1

```
exon      complement(36227. .36865)
          /gene="T1E3_40"
          /number=1
gene      complement(join(36227. .36865,36944. .37374,37440. .37631,
37768. .38448,38589. .38763))
          /gene="T1E3_40"
          /gene="T1E3_40"
          /gene="T1E3_40"
          /gene="T1E3_40"
          /number=1
          /number=2
          /gene="T1E3_40"
          /number=2
          /number=3
          /gene="T1E3_40"
          /number=3
          /gene="T1E3_40"
          /number=4
          /number=5
          /gene="T1E3_50"
          /gene="T1E3_50"
          /number=1
          /number=1

Query Match      56.7%; Score 3895.2; DB 8; Length 83485;
Best Local Similarity 91.3%; Pred. No. 0;
Matches 4334; Conservative 0; Mismatches 13; Indels 401; Gaps 3;

Qy 896 gagtaaatgtagagagatcttttgggaggagtgagtgacttcttcttagcaccagt 955
    |||||
Db 1 GAGTAATAGTGAGAGAGATCTTTTGGGAGGAGTGAGTGACTTCTCCTTTTAGCACCAGT 60
Qy 956 tatcagaacacccagcgttaatgtagaccggtcaatggaattttacttcagatgtggg 1015
    |||||
Db 61 TATCAGAAACACCCCGGTAATGTAGCGCGGTCAATGGAAATTTTACTTCAGATGTGGG 120
Qy 1016 tatgtaaatggtcttccaccagagtggaacttctcaagctggtctataatgagtttga 1075
    |||||
Db 121 TATGCTAAATGGTCTTTTACCACAGAGTGGCACTTCTCAAGCTGGCTATATGAGTTGA 180
Qy 1076 attgtagacttggtgaatcctgatacagatgcccttctctcctcaacagctgctgagtgg 1135
    |||||
Db 181 APTGGATGACTTGTGTAATCTCTGATCAGATGCCCTTCTCCTTCAAGGTTCTGCTGAGTGG 240
Qy 1136 tggggatgacttattcaaggttcgtcaatgtagtgatcaaatctattttcagttttttt 1195
    |||||
Db 241 TGGGGATAGCTTATTCAGGTTTCGTCATGTGAGTGATCAAAATCTATTTCAGTTTCTTTT 300
Qy 1196 ttttcccttcttcggttcttgcagtagcttagagtagaacaatgaaatcttcttaag 1255
    |||||
Db 301 TTTTCCCTTTCTTCGTTCTTCCAGTACTTAGAGTAGAACATGAATTAGATATCTTAAG 360
Qy 1256 aaagtcaatggttttgaacagatggaacctccagcgtgtaacaaagcctctttcaaatgttaa 1315
    |||||
Db 361 AAAGTCATGGTTTTGAACAGATGGACCTCCAGCGGTGAACAGCGCTCTTTTACAATTTGAA 420
Qy 1316 ttccacaattagaacagagagttgggtcagctcgtgaaagttcgtttcaatatgtacc 1375
    |||||
Db 421 TTCACCAATTAGAAGAGAGCGAGTTGGGTGTCAGTCTGTGAAAGTTTCGTTTCAATATGTACC 480
```

```
Qy 1376 gtcaacgcccagctcgtttcagaaacaggtgaaagactggttcttcttgaacagatagttac 1435
    |||||
Db 481 GTCAACGCCAGCTGCTTTCAGAACAGGTGAAAGACTGGATTCTTGAACATAGTTAC 540
Qy 1436 aactactggacaatgaaa tccacagagccgaaatctctgacaaaagtatgcagagcattatgga 1495
    |||||
Db 541 AACTACTGGACATGAAATCCCGAGAGCCGAAATCTCACAAAAGTATGCAGAGCATTTATGA 600
Qy 1496 ctctcgtctcgttaatgcagaggaagctactgaacaaatgatgcacagaacaagatgt 1555
    |||||
Db 601 CTCGTCTGCTGTTAATGCGACGGAAGCTACTGAAACAAATGATGGCAGACAAGATGT 660
Qy 1556 tctggagttcgaccttaacaaaactcctcagcagaaaacccctccaaaagaaaagaaagt 1615
    |||||
Db 661 TCTGCAGTTCGACCTTAACAAAACCTCTCAGCAGAAAACCTCCAAAAGGAAAGAAAGTT 720
Qy 1616 catgcccagggtggtcgtggaaggcaaaccttaaaagaaagccacgcaaacctgcagaact 1675
    |||||
Db 721 CATGCCAAGGTGGTCTGTGGAAGGCAAACTTAAAAAGAAAGCCACGCAAACTTGCAGA 780
Qy 1676 tcccaagtgtcgtggaagcaaaccttaaaagaaagccacgcaaacctgcagaactcagga 1735
    |||||
Db 781 TCCCAAGTGGTGGTGGAGGCAAACTTAAAGGAAGGCCCAAGCTGCRAACTCAGGA 840
Qy 1736 aaaagtgaatctaaagaaacccggagtgccaaaagaaaaatttgaagaatacagcaac 1795
    |||||
Db 841 AAAAGTGAATCTAAAGAAACCCGGGAGTGCCAAAAGAAAAAATTTTGAAGAATCAGCAAC 900
Qy 1796 taaaaagccagccaatgttgagatatgagcaacaaaagccctgaagtcaactcaaaag 1855
    |||||
Db 901 TAAAAAGCCAGCCAATGTGGAGATATGAGCAACAAAAAGCCCTGAAGTCACACTCAAAAG 960
Qy 1856 ttgcagaaaactttgaaattttgacttgagaatcctggagatcgagcgaagtgactc 1915
    |||||
Db 961 TTGCAGAAAAGCTTTGAAATTTTGACTTGAGAATCCTGGAGATGGAGCAAGGTGACTC 1020
Qy 1916 ttagtctgaaattgtccagaacagtagtggcgcaaacctgttttctgagatcagagatgc 1975
    |||||
Db 1021 TGAGTCTGAAATTTGCCAGAACAGTAGTGGCGCAAACTCGTTTTCTTGAGATCAGAGATGC 1080
Qy 1976 cattggtggaactaatgttagtttctcctgattcagtgatcaaaaatagacaagaacaaatgg 2035
    |||||
Db 1081 CATTTGGTGAACATAATGTAGTTTCTTGGATTCACTGTCACAAATAGACAAACCAATGG 1140
Qy 2036 attggggctatgaaccagccacttgaagtgtcaatggaaacccagccagataaaactatc 2095
    |||||
Db 1141 ATTGGGGCTATGAACACAGCCACTTTGAAGTGTCAATGGGAAACCAACCCAGATATAACTATC 1200
Qy 2096 tacaggagcgaactggccagagaccacaacactgatttattgactagaaaaccagcaatg 2155
    |||||
Db 1201 TACAGGAGCGAAACTGGCCAGAGACCAACAACCTGATTTATTGACTAGAAAACCAACAATG 1260
Qy 2156 ccagttcccaatggcaacccagaaacacccagttcccaatggaaaacccaaagcttgct 2215
    |||||
Db 1261 CCAGTTCCTCAGTGGCAACCCAGAACACCCAGTTCCTCAATGGAAAAACCAACAGCTTGGCT 1320
Qy 2216 tcagatgaaaaaccaacttattggctttccatttggtaaccagcaacctcgcataccat 2275
    |||||
Db 1321 TCAGATGAAAAACCAACTTATTGGCTTTTCCATTGTGAACCAACCAACCTCGCATGACCAT 1380
Qy 2276 aagaacacagcagccttcttggccatgggtaatacaaacactatgtatctgtagaggaaac 2335
    |||||
Db 1381 AAGAAACCAAGCAGCCTTCTTGGCCATGGTAAATCAACAACCTATGTATCTGATAGGAAC 1440
Qy 2336 tccacgacctgaattagtaagtggaaacacagcaactagagggtcccccaaggaaacaagcg 2395
    |||||
Db 1441 TCCACGGCTGCTATTAGTAAAGTGGAAACAGCAACTAGGAGGTCCCAAGGAAACAAGCG 1500
Qy 2396 gcctatattttgaataccagagactgtttaccctcctggaaatacagctatatgtacc 2455
    |||||
Db 1501 GCCTATATTTTGAATACCAGACTTGTTTACCTGCTGGAAATCAAGCTATATGGATCACC 1560
```

Qy 2456 tacagacatgcatacaactgttattgtcaaccggaggaggaacaacatggactactgataaa 2515
Db 1561 TACAGACATGCATCAACTGTGTATGTCAACCGGGAGGCAACAACATGCATGACTACTGATAAA 1620
Qy 2516 aaaccagcaactcggatcattataagagccagcagccttgctactcttgattgacca 2575
Db 1621 AAACCAGCAACCTGGATCATTAATAAGAGCCACAGCCTTGCCTACTTGTGATTGACCA 1680
Qy 2576 gcaacctgcaactccaaaaggtttactcacttgaatcagatggttagctaccagcagtc 2635
Db 1681 GCAACCTGCAACTCCAAAAGGTTTACTCACTTCAATCAGATGGTAGCTACACAGCATGTC 1740
Qy 2636 atcgctgggcttgacactcaatctcagtcacaagttccctacaacatatctacatgtgga 2695
Db 1741 ATCGCCTGGGCTTCGACCTCAATCTCAGTCACAAGTTTCTTACAACATATCTACATGTGGA 1800
Qy 2696 atcgtttccaggatttgaatggactacaggtacatgcagagagcagggcctctgc 2755
Db 1801 ATCTGTTCCAGGATTTTGAATGGGACTACAGGTACATGCCAGAGAACGAGGGCTCTCTGC 1860
Qy 2756 atacgattctttacagcaagatatccatcaaggagaaataagtacatactttctcatgagat 2815
Db 1861 ATACGATTTCTTACAGCAAGATATCCATCAAGGAATAAGTACATACTTTCTCATGAGAT 1920
Qy 2816 atccaatggttaattgggtgcaagaagcgttacctcaaaactctctctgcacactccaat 2875
Db 1921 ATCCAATGTTAATGGGTGCAAGAAGCGTTTACCTCAAAACTCTTCTCTGCCAACTCCAAT 1980
Qy 2876 tatggctaaactcgaagaagccaggctcgaagacagtatcatcgtgcaatgggaca 2935
Db 1981 TATGGCTTAAACTTGAGGAAGCCAGGGCTCGAAGAGACAGTATCATCTGTCGAATGGGACA 2040
Qy 2936 gacggaagaagcatgatctaaacttagctcaacagattgctcaatcaacagatgtggagag 2995
Db 2041 GACGGAAGAAGCATGATCTAAACTTAGCTCAACAGATTGCTCAATCACAAGATGTGGAGAG 2100
Qy 2996 acataacagcagcagctgtgtggaatatattagatgctgcaagaagaaacgaaatccagaa 3055
Db 2101 ACATAAACAGCAGCAGTGTGTGGATATTTAGATGCTGCAAGAAGAAAGCAAAATCCAGAA 2160
Qy 3056 agtaqtccaagaataattgcatggcatccacctcaggttatagaatacagagatgctc 3115
Db 2161 AGTAGTCCAAAGAAATTTGCATGGCATGCCACCTCAGGTTATAGAAATCGAGGATGATCC 2220
Qy 3116 aactgatgggcaagaaggttaaaaaactgcccagcatcagtaaggtgcatctaaagg 3175
Db 2221 AACTGATGGGCAAGAAAGGTAAAAATACTGCCAGCATCAGTAAGGTGCATCTAAGG 2280
Qy 3176 aaactcgtctccagttaaaaagcagcagaaagaggaatgtattgtcccaaaaagccc 3235
Db 2281 AAACCTCGTCTCCAGTTAAAAAGACAGCAGAAAAAGAGAAATGTATTGTCCCAAAAACGCC 2340
Qy 3236 tgcataaaaggttcgagcaggtagaaaaaatcagtaacctccgctcgtcactgcctcaga 3295
Db 2341 TGCATAAAAGGGTCGAGCAGGTAGAAAAAATCAGTACTCCGCTGCTCATGGCTCTCAGA 2400
Qy 3296 gatccagcttggcaacctactcctccaaagcacactttatcaagaagcaagcctaaagg 3355
Db 2401 GATCCAGCTTTGGCAACCTACTCTCTCAAGACACACTTTTATCAGAAGCAGCCTAAGG 2460
Qy 3356 aaaagggaaggtccatcacagattcaggaaagcaag----- 3394
Db 2461 AAAAGGGAGAAAGTCCATACAAAGATTTCAGGAAAGCAAGAGGTAACATAATGATTCTACA 2520
Qy 3395 ----- 3394
Db 2521 ATCTCTGTGATATATNTTTGAGATTTTGTAGTAACTGATGTGTCCAAACCAGCTCTCTTATCA 2580
Qy 3395 -----aggctccatcaggagaaactctctgtcaggattctatttcggaa 3437
Db 2581 CTGTGTTGGTCGTTGTATAGTCCATCAGGAACTTCTGTGTCTCAGGATCTATTTCGGAA 2640
Qy 3438 ataatctcagagatgcaaaaatctgtatctagagacaaagaaagacagagcaaat 3497

Db 2641 ATAAATTACAGGATGCAAAATCTGTATCTAGGACACAAAAGAAAGACAGACCAAAAT 2700
Qy 3498 gcaatgctctttacaaaaggagatgggtgcactgtttccctatgagagcaagaagcgaaaa 3557
Db 2701 GCAATGCTTTGTACAAAGGAGATGGTGCACCTTGTCCCTTATGAGCAAGAAGCGAAAA 2760
Qy 3558 ccaagaccccaaaagtggacattgacgatgaaacaactcgcatatggaacttactgaaaggg 3617
Db 2761 CCAAGACCCAAAAGTTGACATTTGACGATGAACAACCTCCGATATGAACTTACTGATGGG 2820
Qy 3618 aagagagatgaaaaagaggggatgaagagaagataaaaagaagaagatggtgggaa 3677
Db 2821 AAAGGAGATGAANAAGAGGGGATGAAGAGAGGATAAAAGAAGAGAGTGTGGGAA 2880
Qy 3678 gaagaagaagagagcttctccgaggaggcgtgattcctctcatcgtcgtcagtcacactggta 3737
Db 2881 GAAGAAAGAAGAGCTTCCGAGGAAGGCTGATTCTTTCATCGCTCGCATGCACCTGGTA 2940
Qy 3738 ca----- 3739
Db 2941 CAAGGTGAAGATCCACTTCTCTCAACTCCATTTTATTCACACAAATTAGTAGAATA 3000
Qy 3740 ----- 3739
Db 3001 CTCAAAATGATGTTTGTGTGCAAAATTTTAAATTCACATAGTTAACCATGTCAAATAA 3060
Qy 3740 ----- 3739
Db 3061 TATTCATAATGCATCTTGTGAGAACAGGTGTGCATTATTATGGTCACAGCTGAATGGTTTA 3120
Qy 3740 -----aggag 3744
Db 3121 TGTGCCATTATTTCTTTACTGCTATAGATGACCAATTTAAACTTAAACGTTTACAGAG 3180
Qy 3745 atagacgtttttccaatggaaggatcggtgtgtgtgattcggtcattcggtatggaactttcccta 3804
Db 3181 ATAGACGTTTTTTCGCCATGGAAGGGATCGGTGTTGATTTCGGTTCATTGGAGTTTTCTCTTA 3240
Qy 3805 cacgaatgtctcggtacacctttca----- 3830
Db 3241 CACAGATGTCGCGATCACCTTTTCAGGTTATAGTTGCTTTAATAAATTGAGTTCCA 3300
Qy 3831 -----agctcgtcgttcactcctcagct 3854
Db 3301 AAACATAGAAATTAACCCATCGTGGTTTTTACAATGCAGCTCTGCGTTTCATGCTCTAGCT 3360
Qy 3855 gctcgattccctccaaaatttaagcagcagccgagaagatgaaagaaatgttagaagccta 3914
Db 3361 GCTCGATTCCTCCAAAATTAAGCAGCAGCCGAGAGATGAAGGAATGTTAGAACGCGTA 3420
Qy 3915 gttgtgaagatccagaaggtgattctgaacttaacttaaatgaaatccttcctggtggcagaa 3974
Db 3421 GTTGTGAGATCCAGAAGGATGCATCTGAACTTAATGAATTCCTTCGTGGCAGGAA 3480
Qy 3975 aaggttcaacatccatctgcatacgtgaagttctcgtgggttgatagtggaatcaaaagagcag 4034
Db 3481 AAGGTTCAACATCCATCTGACATGGAAGTTTCTGCGGTGATAGTGGATCAAAAGAGCAG 3540
Qy 4035 ctaaggagactttcaaacctcgtgaattgaagatttaatttcttagaagaagatattcaaa 4094
Db 3541 CTAAGGGGACTTCAAACTCTGGAAATGAAGATTTAATTTCTTAGAGAAGAGATATTCAA 3600
Qy 4095 aatttagaagaggaagtattatcatcacaaagtcttttgcctcggcgcatatttcagtcg 4154
Db 3601 AATTTAGAGAGGAGATATTATCATCACAAGATCTTTTGTATCCGGCGATATTTTTCAGTCG 3660
Qy 4155 tgtgggagagttggtatcctgttcatgtttccaaaatgaagcagcagaggtttccctacaacagg 4214
Db 3661 TGTGGAGAGTTGGATCCTCTTCATGTTCCAAATCAGACGACGAGAGTTTTCCTACACACAGG 3720
Qy 4215 tgtgaacacaaaactgctcagtggaacatcacaaactcagtgcaaaactggagagcccaacttg 4274

Db 3721 TGTGAAACAAAATGTCAGTGGACATCACAAATCAGTGCACAACTGGAGCCCAAACTTG 3780
Qy tctgatgaatttcttccaagggaatgagacgcgcatctatatgaaggatctcgttgat 4334
Db 3781 TCTGATGAATTTGCTTCAAGGGAATGAGACCGCATCTATGAGAGATCTGGTGAT 3840
Qy gttcagaacaagaatacacaattgcgtctcagagaacacctgcttcttgaaacaaatg 4394
Db 3841 GTTCAGAAACAAGAACTACAAATGTCGCTCAGAGAAACCTGATCTTGAAACAAATG 3900
Qy aattggaagactctgtctgttttgggtcagccaagaaatgataactaattgccaacaact 4454
Db 3901 AATTGGAAGACTCTGCTGTTTGGTCAGCCAAGAAATGATCTACTAATTGCGCAACACT 3960
Qy ccttcacagactatgagcagtgctgcactcgacagccacatgctactagacatagagat 4514
Db 3961 CTTTCCACAGCTATGACAGCTGTGCGACTCGACGCCACATGCTACTAGACATGAGGAT 4020
Qy ttggaatgcgaagtgaaggccttggttattcttggatgtccatctcaccagagttgac 4574
Db 4021 TTTGGAATGCAAGGTGAAGGCTTGGTTATTCTTGGATGTCCATCTCACCAAGAGTTGAC 4080
Qy agagttaagaacaaaatgtaccacgaggttttttcagacaagtggaagtgttccaaga 4634
Db 4081 AGAGTAAAGAACAAAATGTACCACGCGAGGTTTTCAGACAAAGTGAAGTGTCCAAGA 4140
Qy gaattcacaggtccagatcataccatcaacgcctcatgaattaccaggaatggattgtcc 4694
Db 4141 GAAITCACAGTTCAGATCATACCATCAACGCTCATGAAITACCAGGAATGGATTGTCC 4200
Qy ggttctcacaagcgcgtccaagaacaccagagcagatacccacaataatcaacaagatgag 4754
Db 4201 GGTTCCTCAAGCGCGTCCAAGAACACCAGGACGATACCCCAACATAATCAACAAGATGAG 4260
Qy atgaataagatcccatcttaacaaaacatttttggatctgctcaactcctcgaagaa 4814
Db 4261 ATGAATAAAGCATCCCATTTACAAAAACATTTTGGATCTGCCTCAACTCTCTCGAAGAA 4320
Qy tgccttaacaagacagtcaggtaccacaacagacaatcacggtgctgtctaccagagatg 4874
Db 4321 TGCCTTACAAGACAGTCCAGTACCAACAGAACATCACGCGATGCTGTCTACCGAGAGAT 4380
Qy agaactctgaagacgtgtgttgatccgctcagtaacaattcaagcttacagaacatatg 4934
Db 4381 AGAACTGCTGAAGACGTGGTTGATCCGCTCAGTAACAATTTCAAGCTTACAGAACAATATG 4440
Qy gtcgaatcaaatccagcaataaagacagacgcgagttgtaatacaaggaacaaatgcc 4994
Db 4441 CTCGAATCAAAATTCAGCAATAAAGACGACGCGCAGTTGAATCAAGGAGACAAATGCC 4500
Qy actattttacgagatgaaagggaacgcttgcgtgatgggaaaaagcctacaagccagttg 5054
Db 4501 ACTATTTTACGAGAGATGAAAGGAGCGCTTGCCTGATGGGAAAAAGCCTACAAAGCCAGTGG 4560
Qy gatagcttcagaaaaagatgctggagggggaatgaaggagacagggaaacgaaaaaacaat 5114
Db 4561 GATAGCTCTCAGAAAAAGATGTGGAGGGGAATGAAGGGAGACAGGAACGAAACAAAACAAAT 4620
Qy atgaattccatagactatgaagcaataagacgtgtagtatcagcgagatttctcaggct 5174
Db 4621 ATGGATTCCATAGACTATGAGAGCAATAAGACGTGCTAGTATCAGCGAGATTTCTGAGGCT 4680
Qy atcaaggaaaggagatgaataacatgttggccgtacgaataaaggatttctcctagaacgg 5234
Db 4681 ATCAAGGAAGAGGGATGAATAACATGTTGGCCGTACGAAATTAAGGTAAATCTACTAATT 4740
Qy atagttaa 5242
Db 4741 TCAGTTGA 4748

RESULT 3
ATSULFRED

LOCUS ATSULFRED 6055 bp DNA linear PLN 17-JUL-1998
DEFINITION A.thaliana gene encoding sulfite reductase.
ACCESSION Y10157
VERSION Y10157.1 GI:2584720
KEYWORDS sulfite reductase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE 1 (bases 1 to 6055)
AUTHORS Bork,C., Schwenh,J.D. and Hell,R.
TITLE Isolation and characterization of a gene for assimilatory sulfite
reductase from Arabidopsis thaliana
JOURNAL Gene 212 (1), 147-153 (1998)
MEDLINE 98326320
REFERENCE 2 (bases 1 to 6055)
AUTHORS Hell,R.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-1996) R. Hell, Lehrstuhl Fuer
pflanzensphysiologie, Ruhr-Universitaet Bochum, Universitaetsstr.
150, 44801 Bochum, D-44801, FRG
FEATURES
source
I. .6055
/organism="Arabidopsis thaliana"
/db_species="Columbia"
/db_xref="taxon:3702"
1..2269
2542..2733
/number=1
join(2542..2733,3092..3864,4004..4299,4396..4592,
4679..4815,4899..5015,5108..5265,5383..5435)
/codon_start=1
/product="sulfite reductase"
/protein_id="CAA71239.1"
/db_xref="GI:2584721"
/db_xref="SPTREMBL:O23650"
/translation="MSSTFRAPAGAATVFTADQKIRGLDALRSSHSVFLGRYGRGG
VPPPPASSSSSPIOAVSTKPTATKRSKEIIEKSNFIRVPLNEELLTEAPNVNE
SAVOLTKFHGSYOQYINREERGSRYSFMLRTKPSKVPNQLYLTMDLADERIGITL
RLTRTFQLHGVKQNLKNTVMSIINKMGSTLGCAGDLNRNLAPAPVYKDYLFPA
QETADNIAALLSPQSGFYIDMWVDGQFMTAEPEVVVKARNDSHGNTFVDSPEPIYG
TQFLPRKRVAVTPTDINSVDLLNDIGVVVSDENGEPOGFIYVGGMGRTHRMES
TFARLAEPIGYVPKEDILYAVKAIYVTQREHRRDRKYSRMKYLIISSWGIERFDVV
EQYQKKFPESELPEWFKSYLGHWEQDGCWFCGLHVDSGRVGIMKKTLREVLEK
YKIDVRITPNQIVLCDDIKTEWKRPIITVLAQAGLLQPEFVDPNLTAMACPAFLPCP
LAITEARGIPSIILKRVANFEKVLGDYDESVIVRTGCPNGCRRPYMAELGLVGDP
NSTQVWLGGTPNLTQTARSPFMDKVKVHDLEKCEPLFYHWKLERQTKESFETTRMG
FEKLKELIDTYKGVSQ"
2734..3091
/number=1
3092..3864
/number=2
3865..4003
/number=2
4004..4299
/number=3
4300..4395
/number=3
4396..4592
/number=4
4593..4678
/number=4
4679..4815
/number=5
4816..4898
/number=5
4899..5015
/number=6
5016..5107
/number=6
5108..5265
/number=7

YFGYSTDDAVVYLMVLNDLIHGLYPEAIVVGVEDVSGMPAFVCPVEDGGVGFYRLHMH
AVADKWTLELLKKEDWQVDITFTLNRWGEKCVVYAEHQALVGHDKTIAFWLMD
KDMYDFMAYDROATPRVDGIALHKMTLRLTJMLGGEGYLNEMWEGFHEWIDFRT
DOHLPGRVLAGNSYDKCRRRFDLGDABEYLYHGLQEDRAMQNLFEYFGWTSEH
QYISRKDGDRVIVFERGNLFLVFNFWHTNSYSDYRIGCSVPKGYKIVLDSNLSLFGG
FNRLLDSAEFTSDGRHDDRPCBFMYAPCORTAVVYAANDDDDDDDSSRLVPLGLLPE
DV"

repeat_region
complement(19524..19568)
/rpt_family="(TA)n"
complement(join(22710..23271,24081..24435,24650..24841,
24940..25347))
/gene="At2g36400"
complement(22710..25347)
/gene="At2g36400"
/note="Fl011.3"
CDS
complement(join(22836..23271,24081..24435,24650..24841,

Query Match 1.7%; Score 117.8; DB 8; Length 101365;
Best Local Similarity 61.4%; Pred. No. 3.6e-16;
Matches 224; Conservative 0; Mismatches 137; Indels 4; Gaps 2;

QY 5802 ggagcaccatcgaaatagagaaaaactgtgaaccataattgaagagcgccctcgcccggg 5861

Db 62343 GAACCAGCAAAAAGGTGCATGTGTGAACCAATCATCGAAGAGCCTGCTTCACCGGAG 62284

QY 5862 caagagtgcactgaataacccgagagtgtattgaagatgcttactacaatgagagacct 5921

Db 62283 CCAGAAACCGAAGATATCAATAGCTGACATAGAGGAGCGGTTTTT---TGAGGATCCA 62227

QY 5922 gacgagatcccaacaataaaactcaacatgaacagtttggaatgaactctacggaacac 5981

Db 62226 GAAGAAATTCCTACCATCAGCTAAACATGATGTCATTTACCACTAAGTGAAGAGATA 62167

QY 5982 atgaagaagaacatggagctccagaaggtgacatgtccaggcttgggttggcttgcac 6041

Db 62166 ATGGAACACACAGGAACCTTCAGACGGAACATGTCCAGCGCTTTAGTTGCACCTACT 62107

QY 6042 ccaacaactctattcccaactcccaactcaaaactaaagaacattagccgtctcaggacagag 6101

Db 62106 GCTGAAACTGCTTCTCTCCATGCCTAAGCTCAAGATATACGCCAGTTAAGGACAGAA 62047

QY 6102 caccagtgacggctcccgatccacatcgctctccttgatgtagtataaaagagaa 6161

Db 62046 CACCGAGTGAAGA-ATCCATTTTCTGTGTTTTTCTGTTCTTCTGTTCTTAGGTTAAAAATGAA 61988

QY 6162 ccaga 6166

Db 61987 TCAGA 61983

RESULT 6
AP002745 97366 bp DNA linear PLN 11-JUL-2000
LOCUS Oryza sativa genomic DNA, chromosome 1, PAC clone:P0489G09,
DEFINITION complete sequence.
ACCESSION AP002745
VERSION AP002745.1 GI:9049478
KEYWORDS HTG.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0489G09.
ORGANISM Oryza sativa

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0489G09

2 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0489G09

2 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0489G09

2 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0489G09

2 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0489G09

2 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0489G09

Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The orientation of the sequence is from SP6 to 77 of the PAC clone.

FEATURES
Location/Qualifiers
source 1..97366
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0489G09"
BASE COUNT 28305 a 20201 c 20553 g 28307 t
ORIGIN

Query Match 1.5%; Score 100.4; DB 8; Length 97366;
Best Local Similarity 55.6%; Pred. No. 3.3e-12;
Matches 257; Conservative 0; Mismatches 196; Indels 9; Gaps 3;

QY 5663 tagtgcagaactgtctttaccggcaccagagagagagcttaacaaagtgcactattcc 5722

Db 17904 TAGTGCAGAGCTCGCTCTTCTTGACCTGAAGAGAAGAGTTTAGTTACATCTGGAACCCC 17963

QY 5723 ggtccctcccgagtccttctcctctgtagccatcccgatgatagaactacctcttcggt 5782

Db 17964 AATAGCTGCAGAAACCTTCCACCAGACATATATAAGTTCTTAGGCCTGTAGTAAGTCAGCT 18023

QY 5783 ggagaaatccctagcaagtggagcaccatcgatagagaaactgtgaaccaataatga 5842

Db 18024 TGA----TGGAAATTCAACACCTGTCACCATGTTATGAACAATGCCAGCAATCATTTGA 18080

QY 5843 agagccggcctcgccgggc---aagagtgcactgaataaccgagagtgatattgaaga 5899

Db 18081 GGAGCCAGCAAGCCCAACCTGGAACATGAGACAGACAGATGAAAGAGTGTGCAATAGA 18140

QY 5900 tgcttactcaatgagagccctgacgagatcccaacaataaaactcaacatttgacaggt 5959

Db 18141 GGATAGTTTTGTGTCATGATCAGAAAGAAATCCCTACTATCAAGCTTAATTTTGAGGAGTT 18200

QY 5960 tgggaatgactctacgggaacacatggaaaga---aacatgagctccaagaagtgcacat 6016

Db 18201 TACACAGAACTGGAAGAGTTATATGCAAGCAAAATPACATGAGATTGAAGATGCTGATAT 18260

QY 6017 gtccagagcttgggtgtcttgatcccaacaactactcttattccaaactcccaactaaa 6076

Db 18261 GTCAAGGCTTGTGTCGCTATACTCTGAAGTTGCTTCTATCCCACTCTTAGCTCAA 18320

QY 6077 gaacattagccgtctcaggacagaccacaaagtgtacagct 6118

Db 18321 GAATGTCAGTCGCCTAAGGACAGACACCACCAAGTGATGATCT 18362

RESULT 7
AP002861 134967 bp DNA linear PLN 26-JAN-2001
LOCUS Oryza sativa genomic DNA, chromosome 1, PAC clone:P0665D10.
DEFINITION complete sequence.
ACCESSION AP002861
VERSION AP002861.2 GI:10800055
KEYWORDS HTG.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0665D10.
ORGANISM Oryza sativa

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10

2 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10

2 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10

2 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10

2 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10

2 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10

Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@ab.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
On Oct 12, 2000 this sequence version replaced gi:9967270.
Genes were predicted from the integrated results of the following: GENSCAN1.0, BLAST2.0, BLASTX2.0 as well as SplicePredictor (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.
The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0665D10 clone has an overlap with P0515G01 (DBJ: AP001633) clone at the position 1 to 26401 of 5' end. The sequence of this clone starts at the position 148672 of P515G01. This sequence of P0665D10 clone has an overlap with P0489G09 (DBJ: AP002745) clone at the position 54916 to 134967 of 3' end. The sequence of this clone ends at the position 80052 of P0489G09. Detailed information on overlap and assembly quality together with annotation of this entry is available at
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES

source
1. 134967
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0665D10"
join(4426..4444,6273..6352,6399..6564,7311..7426)
/gene="P0665D10.1"
join(4426..4444,6273..6352,6399..6564,7311..7426)
/gene="P0665D10.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6476.1"
/db_xref="GI:10800056"
/translation="MSIHESGRAYGIPPIKLFVIVVFVKLICSPKVELCAKIRSQYT
TVVLSWISLTENEIGFLLRTR"
complement(8189..8695)
/gene="P0665D10.2"
complement(8189..8695)
/gene="P0665D10.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6477.1"
/db_xref="GI:10800057"
/translation="MNLPPSKETAEDPSPRPNSLSPSPQRPMLHRSRSRAR
GASMAITGTPSLTCPFPSPSTARVNSPSSHVPAAADVAAVAVSLPPK
PSRSPYPYGNRDPDIANSPTSPRSAPRRRRPAGGQLELSPTAATMAIPAENVL
EKPPFTSV"
complement(join(9395..9695,10080..10238,10320..10495,
10711..10814,11499..11616,11688..11871,11960..12270))
/gene="P0665D10.3"
complement(join(9395..9695,10080..10238,10320..10495,
10711..10814,11499..11616,11688..11871,11960..12270))
/gene="P0665D10.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6478.1"
/db_xref="GI:10800058"
/translation="MRGRAAVFAVLVFLVACTAAAAAAAITSRKHOHRPASGAA
KAACDPAAGSVVVDSEYPIYDSATCFPIRAEFCRIGRPDKKEYLKRYRQPPCST

PRFDGVALLRMSGKKVMFVGDSLALNQYESLLCMLHAAAPNARTTVPASGKVDPLT
TARFEPWRLTMMSSSEPTVTITTTSCILVRREPLASRHDHGFARRALGRYFAQRTYAR
RRARDACIFGCAAVFSTVLEFNVTIYLYLVIVNEKAGRLVKLDAIDQARNW
LSADLVFDSWHMWRSGPTQPDYIOEGNTVVKDMRTEAFSAKALNTWAWWDANLL
QNTRVFFQGISPSHYRGDWDGTASATMGOTFPYNGTAYPGPIPOQAFYALALAGH
" join(14847..15053,15199..15396,15690..15863,15919..16010,
16157..16269,16388..16812)
/gene="P0665D10.4"
join(14847..15053,15199..15396,15690..15863,15919..16010,
16157..16269,16388..16812)
/gene="P0665D10.4"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6479.1"
/db_xref="GI:10800059"
/translation="MAYKLYGALTCLEHRRKKVPDYCEKTKLSMKFIRLIKQCNSSSD
KEFSPLSKRLSKSTLVDEGFAPFWDIVLKPLFFHGRLPDPIDNSNVAVIICQSVT
VTRASGVSSDDDELAPPTPHPKPAGVVRPSMARVGEQLHQQAPAGAGGGGRVFA
ARGAARRAQAAGSFMGDDTDCGGGGGRLVALDAGGEGEDGDEVLAIIRPREGR
RRSARLPLLLDLTWTSTRILLFILNLTMSNAEIOLOITGVRGEQDVVAESDKG
TSRADATAAAAAPPVSSKRLRWAVLANIVFVLGGQSGVCFESAGYTIRIGDSLIW
LATVQSCGAPLTVLILLLLYFQMGSPAPPLIEPSEKRWQSSAGAGLDGLRSPSAG
SVBKRCGEKK"
complement(join(17545..18164,18260..18689))
/gene="P0665D10.5"
complement(join(17545..18164,18260..18689))
/gene="P0665D10.5"
/codon_start=1
/product="putative cysteine protease"
/protein_id="BABI6480.1"
/db_xref="GI:10800060"
/translation="MAPSRPVAIAVILLCTFLAQAADAAYGGGDDGYTOMFEEM
MAFKGKYPCHGEKRYFRGFRNVRFRIRTPAGNLSALRVNQFADLTNDFEVSTH
TGAKPPCKDAPRGVDPDWLPCCIDWRYKGAVTVDVQDQACGSCWAEVAAALEGLTQ
IRTKLTLPLSEQLVDCDSCGCGAGTDFRAELVAAGKGTAEAGRYEYGRKGR
ADALFNHAARIGHRVAVPGDQRLATAVARQPVTAIDASGPAFOFYGSGVPPGPC
GSGSGAAAAPTTNHATVLYGYCQDQASGKKYVAKNSWCKTWGCKYILLEKDVASP
HGTGCVAVSPFYPTV"
join(22285..22792,22869..23461)
/gene="P0665D10.6"
join(22285..22792,22869..23461)
/gene="P0665D10.6"
/codon_start=1
/product="putative cysteine protease"
/protein_id="BABI6481.1"
/db_xref="GI:10800061"
/translation="MASSKPSYLSLFSIACLSLPLSAANPSPPPPCKSKDKE
MRLFSRWMAQYKGYKASWPIEHKRYQIMKDNSNFIGSFRSETEISSGVGAFAQVTV
DSFVGMNRFCDLTPGEFAEQFTGNATGGLLHAAPPCPIPPDSMLPCCVDMRSSGAV
TGKVFORSASCASNAFAAAALEGLNKIRTGELVSLSEQVWVDCDTSGSGSGGRADTA
LGLVAARGGVASSEEPYTGVRGCDGVKLSGHSASLSGFRVAVPPNDERQLALAVAR
QPVTAYIDAGAREFMFYKGGVYRGPCEAERNHAVAIVGYCEGFGDKGYIAKNSWGS
DWGEQGYVYLAKDVWMPQGTGLATSPFYPTA"
complement(join(24388..24528,24577..24696,24937..25103,
25260..25361,25832..25972,26080..26163,26253..26559,
26861..26727,27036..27168,27254..27347,27356..27496,
28444..28512,29291..29499,29598..29792,30489..30538,
30822..31230,31307..31652,32161..32267))
/gene="P0665D10.7"
complement(join(24388..24528,24577..24696,24937..25103,
25260..25361,25832..25972,26080..26163,26253..26559,
26861..26727,27036..27168,27254..27347,27356..27496,
28444..28512,29291..29499,29598..29792,30489..30538,
30822..31230,31307..31652,32161..32267))
/gene="P0665D10.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6482.1"
/db_xref="GI:10800062"
/translation="MAILEDNDLLSLPGLSANTRIGESPQHSDLLILGRPQLCR
RRRTSLPRLRRNRGLPRLSPHSTATPTEDRAVVAVPGRSHRRPPPHSTAAPRRSGRP


```

TITLE      Direct Submission
JOURNAL    Submitted (26-DEC-2001) Environmental Horticulture, University of
           California, One Shields Avenue, Davis, CA 95616, USA
FEATURES   Location/Qualifiers
           source
             1..160
               /organism="Narcissus pseudonarcissus"
               /cultivar="Dutch Master"
               /db_xref="taxon:39639"
               /tissue_type="4-day old flower tepal"
               <1..>160
               /codon_start=1
               /product="unknown"
               /protein_id="AA169364.1"
               /db_xref="GI:18419584"
               /translation="PSIFKGLTREGIQHCFWGVGVGFDMMRAPKPLFARLHFFA
               SKGRNRT"
BASE COUNT 51 a 32 c 41 g 36 t
ORIGIN
Query Match      1.2%; Score 79.4; DB 8; Length 160;
Best Local Similarity 69.9%; Pred. No. 2.9e-07;
Matches 107; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 6511 catcaattcagaggttttcaacgagcagatcacagtctgttcttggaagatttcg 6570
Db 2 CATCAATATTTAAAGCCTTACAACTGAAGGATACAGCATTGCTTTTGGAGAGGTTTG 61
Qy 6571 tatgttcgtgattcgaacagaagaagacagaccgcgtccatttaattgcaaggttgc 6630
Db 62 TGTGCGTGAAGGGATTCGATCGGATGATGAGAGCACCACCAACCTCTCTTGCAGGCTTC 121
Qy 6631 atttcttcgagcgaattgaagaacacacaaa 6663
Db 122 ACTTCCAGCAGCAAGAGNAGTACGACACAAA 154

RESULT 10
LOCUS      I66494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION   I66494.1 GI:2724471
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE  1 (bases 1 to 7218)
AUTHORS   Dörner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE     Recombinant fowlpox virus
JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES   Location/Qualifiers
           source
             1..7218
               /organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match      1.1%; Score 73; DB 6; Length 7218;
Best Local Similarity 3.5%; Pred. No. 6.7e-06;
Matches 13; Conservative 227; Mismatches 127; Indels 0; Gaps 0;
Qy 50 ctttagctcattctctgttatttacattattcctttttctctctccctttt 109
Db 1091 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1150
Qy 110 tatctggaattgttctgctaaaatttccagctgttacattttccgatacagagaat 169
Db 1151 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1210
Qy 170 cactgggttttatgttaatacaatcatgcttctgtttcttgatcataaatacgtat 229
Db 1211 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1270
```

```

Qy 230 taacacctgatttgattctgcgtaataaaaaacctgattgttcttttatctcatttc 289
Db 1271 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1330
Qy 290 ccataaacattgcttacttattcttcgtctcttcttaccgtttccagctaaaaattctt 349
Db 1331 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1390
Qy 350 cgcattcaatgcttctcgttcttcttgatgagaaaaaatatctgcacaaaatcattta 409
Db 1391 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1450
Qy 410 ttgcatt 416
Db 1451 TCTCTTT 1457

RESULT 11
LOCUS      CNS06EVQ
DEFINITION T7 end of clone AR0AA025B09 of library AR0AA from strain CBS 732 of
           Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION  AL395628
VERSION     AL395628.1 GI:12147375
KEYWORDS   STS.
SOURCE     Zygosaccharomyces rouxii.
ORGANISM   Zygosaccharomyces rouxii.
REFERENCE  1 (bases 1 to 863)
AUTHORS    de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B.,
           Wincker,P., Artiguenave,F. and Souciet,J.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 8.
           Zygosaccharomyces rouxii(1)
JOURNAL    FEBS Lett. 487 (1), 52-55 (2000)
PUBMED     11152883
REFERENCE  2 (bases 1 to 863)
AUTHORS    Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
           Bolotin-Fukuhara,M., Bon,E., Bröttier,P., Casaregola,S.,
           de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
           Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
           Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
           Wincker,P. and Weissenbach,J.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
           yeast species for molecular evolution studies(1)
JOURNAL    FEBS Lett. 487 (1), 3-12 (2000)
PUBMED     11152876
REFERENCE  3 (bases 1 to 863)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
           seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
           This STS is part of a random genomic sequencing program of thirteen
           yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
           exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
           Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
           lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
           angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
           Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
           5 kb were prepared and both extremities were sequenced. See
           keywords for description of this sequence and for the sequence of
           the other extremity of this insert.
FEATURES   Location/Qualifiers
           source
             1..863
               /organism="Zygosaccharomyces rouxii"
               /strain="CBS 732"
               /db_xref="taxon:4956"
               /clone="AR0AA025B09"
               /clone_lib="AR0AA"
               /note="end : T7"
           misc_feature
             <3..>155
```

/note="similar to Saccharomyces cerevisiae ORF YGR184c [UBR1 ; ubiquitin-protein ligase]"
/evidence=not_experimental

BASE COUNT 196 a 105 c 60 g 417 t 85 others
ORIGIN

Query Match 1.08; Score 66.2; DB 11; Length 863;
Best Local Similarity 42.3%; Pred. No. 0.00027;
Matches 157; Conservative 40; Mismatches 173; Indels 1; Gaps 1;

Qy 51 ttttagtcacatttctgttatttattacatatatttcctttttttctctctctccctttttt 110
Db 255 TTCTATTATTTATATATATATATTTTATTKTATGTCCTTTTCTTTTCTTTTCTTTT 314
Qy 111 atctgggaattgtctgtaaaattccagctgtgtacatttccgatacagagaagaatc 170
Db 315 WTTYCYCTWTTTTTAACTTYYTTTCTTCYTYAATTTTTTWTATYACAAACCTT 374
Qy 171 actgggttttatgtaataacatagttctctctgttttctgataataatctcagctatt 230
Db 375 CTCTCTTTWTTTWTWTTTATATTTTATTTTATTTTCTCTTWCCTTTATTTAYTTTTT 434
Qy 231 aacacctgatttgcattcggtcgtaaaaaaacctctgatttgccttttattcttcacttcc 290
Db 435 YCCAYTCAWTTTTTCT 494
Qy 291 ccataaacattgctacttattcctctctctctctctctctctctctctctctctctct 349
Db 495 TTACTATTCTCTTATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 554
Qy 350 cgtattcaatggttctcgtctgtttgttgatgagaaaaatctgcacaaaaaatcattta 409
Db 555 AWTTTTTTATTTGTTTCTTTAAWTTTWTACTTWTCTTCTCTCTCTCTCTCTCTCTCT 614
Qy 410 ttgcattttat 420
Db 615 TATTATTTTTT 625

RESULT 12
AC073669/c
LOCUS AC073669 152968 bp DNA linear HTG 29-JUN-2000
DEFINITION Mus musculus clone CT7-273D11, WORKING DRAFT SEQUENCE, 39 unordered
pieces.

ACCESSION AC073669
VERSION AC073669.1 GI:8810286
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 152968)
AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Mouse
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 152968)
AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1058096
Center clone name: RG-MBAC_273D11

Summary Statistics
Consensus quality: 118135 bases at least Q40
Consensus quality: 134236 bases at least Q30

Consensus quality: 137982 bases at least Q20
Estimated insert size: 100000; pulse field gel estimation
Estimated insert size: 149168; sum-of-contrigs estimation
Quality coverage: 7.41 in Q20 bases; pulse field gel estimation
Quality coverage: 4.96 in Q20 bases; sum-of-contrigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1093: contig of 1093 bp in length
* 1094 1193: gap of unknown length
* 1194 2195: contig of 1002 bp in length
* 2196 2295: gap of unknown length
* 2296 3335: contig of 1040 bp in length
* 3336 3435: gap of unknown length
* 3436 4873: contig of 1438 bp in length
* 4874 4973: gap of unknown length
* 4974 6084: contig of 1111 bp in length
* 6085 6184: gap of unknown length
* 6185 7742: contig of 1558 bp in length
* 7743 7842: gap of unknown length
* 7843 8945: contig of 1103 bp in length
* 8946 9045: gap of unknown length
* 9046 10120: contig of 1075 bp in length
* 10121 10220: gap of unknown length
* 10221 11475: contig of 1255 bp in length
* 11476 11575: gap of unknown length
* 11576 12699: contig of 1124 bp in length
* 12700 12799: gap of unknown length
* 12800 13987: contig of 1188 bp in length
* 13988 14087: gap of unknown length
* 14088 15182: contig of 1095 bp in length
* 15183 15282: gap of unknown length
* 15283 16467: contig of 1185 bp in length
* 16468 16567: gap of unknown length
* 16568 18435: contig of 1868 bp in length
* 18436 18535: gap of unknown length
* 18536 20772: contig of 2237 bp in length
* 20773 20872: gap of unknown length
* 20873 22213: contig of 1341 bp in length
* 22214 22313: gap of unknown length
* 22314 23495: contig of 1182 bp in length
* 23496 23595: gap of unknown length
* 23596 25217: contig of 1622 bp in length
* 25218 25317: gap of unknown length
* 25318 27409: contig of 2092 bp in length
* 27410 27509: gap of unknown length
* 27510 28865: contig of 1356 bp in length
* 28866 28965: gap of unknown length
* 28966 31579: contig of 2614 bp in length
* 31580 31679: gap of unknown length
* 31680 34519: contig of 2740 bp in length
* 34520 38002: contig of 3483 bp in length
* 38003 38102: gap of unknown length
* 38103 42293: contig of 4191 bp in length
* 42294 42393: gap of unknown length
* 42394 47356: contig of 4963 bp in length
* 47357 47456: gap of unknown length
* 47457 51699: contig of 4243 bp in length
* 51700 51799: gap of unknown length
* 51800 55044: contig of 3245 bp in length
* 55045 55144: gap of unknown length
* 55145 59713: contig of 4569 bp in length
* 59714 59813: gap of unknown length
* 59814 64897: contig of 5084 bp in length
* 64898 64997: gap of unknown length
* 64998 69683: contig of 4686 bp in length
* 69684 69783: gap of unknown length

[illegible]

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2002, 05:51:56 ; Search time 650.38 Seconds
 (without alignments)
 18143.782 Million cell updates/sec

Title: US-09-840-743-5
 Perfect score: 6873
 Sequence: 1 gtctccgcatgtactcgc.....aaaaaaaaaaaaaacctcgag 6873

Scoring table:
 IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6873	100.0	6873	24	AAS96689 Arabidopsis cDNA e
2	4790.2	69.7	10620	24	AAS96691 Arabidopsis DMT1 (
3	4788.6	69.7	12785	24	AAS96686 Arabidopsis DNA fo
4	1478	21.5	1478	24	AAS96690 Arabidopsis DMT 5'
5	411.8	6.0	758	24	AAS96707 Tomato DMT related
6	352.6	5.1	638	24	AAS96702 Soybean DMT1 DNA.
7	351.6	5.1	657	24	AAS96706 Alfalfa DMT relate
8	338.8	4.9	765	24	AAS96710 Cotton DMT related
9	329	4.8	1134	24	AAS96703 Soybean DMT2 DNA.

10	253	3.7	616	24	AAS96704 Soybean DMT3 DNA.
11	233	3.4	798	24	AAS96696 Corn DMT1 DNA. 2e
12	228.2	3.3	517	24	AAS96709 Corn DMT related c
13	219.6	3.2	583	24	AAS96711 Soybean DMT relate
14	205	3.0	205	24	AAS96688 Arabidopsis DMT ge
15	191.8	2.8	595	24	AAS96714 Corn DMT related c
16	179.8	2.6	640	24	AAS96715 Corn DMT related c
17	154	2.2	557	24	AAS96698 Wheat DMT1 DNA. T
18	151	2.2	439	24	AAS96699 Wheat DMT2 DNA. T
19	144.6	2.1	488	24	AAS96718 Corn DMT related c
20	142.4	2.1	422	24	AAS96717 Corn DMT2 DNA. 2e
21	135.6	2.0	506	24	AAS96697 Soybean DMT4 DNA.
22	131.8	1.9	457	24	AAS96705 Wheat DMT3 DNA. T
23	121.6	1.8	624	24	AAS96700 Corn DMT related c
24	118	1.7	570	24	AAS96719 Arabidopsis DMT2 (
25	117.8	1.7	8880	24	AAS96692 Corn DMT related c
26	116.4	1.7	613	24	AAS96720 Corn DMT related c
27	108.4	1.6	561	24	AAS96716 Wheat DMT4 DNA. T
28	100.4	1.5	492	24	AAS96701 Rice DMT1 DNA. Or
29	100.4	1.5	12120	24	AAS96695 Tomato DMT related
30	68.4	1.0	589	24	AAS96713 Tumour suppressor
31	65	0.9	5750	22	AAS46708 Human immune syste
32	65	0.9	5750	24	ABL34008 Arabidopsis DMT3 (
33	63.2	0.9	8760	24	AAS96693 Human immune syste
34	60.2	0.9	16033	24	ABL33404 Human immune syste
35	58.8	0.9	6145	24	ABL32972 Human immune syste
36	58.4	0.8	14006	24	ABL33958 Human immune syste
37	58	0.8	6419	24	ABL32267 Human immune syste
38	57	0.8	6668	24	ABL33696 Tumour suppressor
39	55.8	0.8	7442	22	AAS46686 Arabidopsis DMT4 (
40	55.8	0.8	8580	24	AAS96694 Human immune syste
41	55.4	0.8	12138	24	ABL33628 Base sequence of t
42	53.8	0.8	6644	20	AAX33181 Base sequence of t
43	53.8	0.8	7372	20	AAX33182 Base sequence of t
44	53.8	0.8	7797	20	AAX33180 Cowpox virus bsr f
45	53.8	0.8	7996	20	AAX33184 Base sequence of t

ALIGNMENTS

RESULT 1

AAS96689 ID AAS96689 standard; cDNA; 6873 BP.

XX AC AAS96689;

XX DT 26-FEB-2002 (first entry)

XX DE Arabidopsis cDNA encoding Demeter protein.

XX DEMETER; DMT; Atropis; ATR; 5-methylcytosine glycosylase; ss;

XX DNA demethylation; transgenic plant; transcription modulation;

KW flowering time; endosperm development; MEDEA.

XX OS Arabidopsis Thaliana.

XX PN WO200180626-A1.

XX PD 01-NOV-2001.

XX PF 23-APR-2001; 2001WO-USI3059.

XX PR 21-APR-2000; 2000US-0553690.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;

XX DR WPI; 2002-055307/07.

XX DR P-PSDB; AAU72736.

XX PT New polynucleotide that control plant development comprising a sequence

Db 1621 ccaaggtggtcgtggaggcgaacctaaagaaagaaagccacgcgcaaacctgcagaaacttccca 1680
Qy 1681 aagtggtcgtggaggcaaacctaaagaaagaaagccacgcgcaaacctgcagaaactcagaaaaag 1740
Db 1681 aagtggtcgtggaggcaaacctaaagaaagaaagccacgcgcaaacctgcagaaaaag 1740
Qy 1741 tgaatctaaagaaacccgggagtgccaaaaagaaaaatttgaagaaatcagcaactaaaa 1800
Db 1741 tgaatctaaagaaacccgggagtgccaaaaagaaaaatttgaagaaatcagcaactaaaa 1800
Qy 1801 agccagccaatggtggagatagagcaacaaagccctgaaagtccacactcaaaaagtgc 1860
Db 1801 agccagccaatggtggagatagagcaacaaagccctgaaagtccacactcaaaaagtgc 1860
Qy 1861 gaaaagcttgaatttgacttgagaatcctggagatgcgagagcaagtgactctgagt 1920
Db 1861 gaaaagcttgaatttgacttgagaatcctggagatgcgagagcaagtgactctgagt 1920
Qy 1921 ctgaaattgtccagaacagtagtggcgcaaacctcgtttcttgagatcagagatgccattg 1980
Db 1921 ctgaaattgtccagaacagtagtggcgcaaacctcgtttcttgagatcagagatgccattg 1980
Qy 1981 gtgaaactaatgttagtttctcgtgattcagtgtcacaaatagacaagaccaatgattgg 2040
Db 1981 gtgaaactaatgttagtttctcgtgattcagtgtcacaaatagacaagaccaatgattgg 2040
Qy 2041 gggctatgaacccagccacttgaagtgtcaatgggaaacccagccagataaactatctacag 2100
Db 2041 gggctatgaacccagccacttgaagtgtcaatgggaaacccagccagataaactatctacag 2100
Qy 2101 gagcgaaactggccagagaccacaacactgatttattgactagaaacccagcaatggccagt 2160
Db 2101 gagcgaaactggccagagaccacaacactgatttattgactagaaacccagcaatggccagt 2160
Qy 2161 tccagtggcgaacccagaaacccagttcccaatggaaacccaaacaaagcttgggttcaga 2220
Db 2161 tccagtggcgaacccagaaacccagttcccaatggaaacccaaacaaagcttgggttcaga 2220
Qy 2221 tgaaaaccaaacttattggctttccatttggtaaccagcaacctcgatgagaaactccac 2280
Db 2221 tgaaaaccaaacttattggctttccatttggtaaccagcaacctcgatgagaaactccac 2280
Qy 2281 accagcagccttgggtggccatgggtaatacaaacctatgtatctgatatgaaactccac 2340
Db 2281 accagcagccttgggtggccatgggtaatacaaacctatgtatctgatatgaaactccac 2340
Qy 2341 ggcctgcatagtagtggaacccagcaactagaggtccccaaggaacaaagcgcccta 2400
Db 2341 ggcctgcatagtagtggaacccagcaactagaggtccccaaggaacaaagcgcccta 2400
Qy 2401 tattttgaatcacagactgtttacctgctggaatcagctatatgtgacactcacag 2460
Db 2401 tattttgaatcacagactgtttacctgctggaatcagctatatgtgacactcacag 2460
Qy 2461 acatgcatcaactgttatgtcaaccggagggaacaaactatgactgactgtataaaaaacc 2520
Db 2461 acatgcatcaactgttatgtcaaccggagggaacaaactatgactgactgtataaaaaacc 2520
Qy 2521 agcaacctggatcataataagagccagcagccttcggtaccctttgattgaccagaac 2580
Db 2521 agcaacctggatcataataagagccagcagccttcggtaccctttgattgaccagaac 2580
Qy 2581 ctgcaactccaaaaggttttactcaacttgaatcagatgggtagctaccagcatgtcatcgc 2640
Db 2581 ctgcaactccaaaaggttttactcaacttgaatcagatgggtagctaccagcatgtcatcgc 2640
Qy 2641 ctgggcttcgacctattctcagtcacaagtctctacaacatacttaoatgtggaatctg 2700
Db 2641 ctgggcttcgacctattctcagtcacaagtctctacaacatacttaoatgtggaatctg 2700
Qy 2701 ttccagagatttgaatgggactacaggtacatgccagagaagcagggctcctctacag 2760
Db 2701 ttccagagatttgaatgggactacaggtacatgccagagaagcagggctcctctacag 2760

Qy 2761 atttttacagcaagatattccatcaaggaataaagtacatacttttctcatgagatatcca 2820
Db 2761 atttttacagcaagatattccatcaaggaataaagtacatacttttctcatgagatatcca 2820
Qy 2821 atggttaattgggtgcaagaaagcgttacctcaaaactcttctctgccaactccaattatgg 2880
Db 2821 atggttaattgggtgcaagaaagcgttacctcaaaactcttctctgccaactccaattatgg 2880
Qy 2881 ctaaaacttgagaaagccaggggctcgaagagacagtagtcatcgtgcaatggagacagcg 2940
Db 2881 ctaaaacttgagaaagccaggggctcgaagagacagtagtcatcgtgcaatggagacagcg 2940
Qy 2941 aaaagcatgattctaaacttagctcaacagattgtctcaatcaacagatgtggagagacata 3000
Db 2941 aaaagcatgattctaaacttagctcaacagattgtctcaatcaacagatgtggagagacata 3000
Qy 3001 acagcagcacgtggtggaatatttagatgctgcaagaaacccagaaatccagaaagttag 3060
Db 3001 acagcagcacgtggtggaatatttagatgctgcaagaaacccagaaatccagaaagttag 3060
Qy 3061 tccaagaaaatttgcattggcagtcacacctgaggttatagaatcgagagatgccaaactg 3120
Db 3061 tccaagaaaatttgcattggcagtcacacctgaggttatagaatcgagagatgccaaactg 3120
Qy 3121 atggggcaagaaaaggttaaaatattgctccagcatcagtaaaaggtgcatctaaagaaaaact 3180
Db 3121 atggggcaagaaaaggttaaaatattgctccagcatcagtaaaaggtgcatctaaagaaaaact 3180
Qy 3181 cgtctccagtttaaaagacagcagaaagggaaatgtattgtccccaaaaacgcctgcaa 3240
Db 3181 cgtctccagtttaaaagacagcagaaagggaaatgtattgtccccaaaaacgcctgcaa 3240
Qy 3241 aaaagggctcagacaggttagaaaaaaatcagtaacctcgctcgctcatcctcaagatccc 3300
Db 3241 aaaagggctcagacaggttagaaaaaaatcagtaacctcgctcgctcatcctcaagatccc 3300
Qy 3301 agctttggcaaccttactctccaaagacacacctttatcaagaaagcaagcctaaagaaaaag 3360
Db 3301 agctttggcaaccttactctccaaagacacacctttatcaagaaagcaagcctaaagaaaaag 3360
Qy 3361 ggaagaaagtcacacagattcagaaaaagcaagaggttccatcagggagaactctgtgtc 3420
Db 3361 ggaagaaagtcacacagattcagaaaaagcaagaggttccatcagggagaactctgtgtc 3420
Qy 3421 aggaattctattgcgaaataatttacaggaatgcataatctgtatctatcaggaacaaaaa 3480
Db 3421 aggaattctattgcgaaataatttacaggaatgcataatctgtatctatcaggaacaaaaa 3480
Qy 3481 gagaacaaagcaaaaatgcaatggtcttgcacaaagagatgggtgacctgttccctatg 3540
Db 3481 gagaacaaagcaaaaatgcaatggtcttgcacaaagagatgggtgacctgttccctatg 3540
Qy 3541 agagcaagaaacgaaaaacccaaagacccaaagtgcacattgacgatgaaaaactcgcatat 3600
Db 3541 agagcaagaaacgaaaaacccaaagacccaaagtgcacattgacgatgaaaaactcgcatat 3600
Qy 3601 ggaacttactatgggaaagagatgaaaaaggggagatgaaagaaaggaataaaaaaaga 3660
Db 3601 ggaacttactatgggaaagagatgaaaaaggggagatgaaagaaaggaataaaaaaaga 3660
Qy 3661 aagagaagtgttgggaagaaagaaagagatcttcccgagggaagggctgattcccttcacg 3720
Db 3661 aagagaagtgttgggaagaaagaaagagatcttcccgagggaagggctgattcccttcacg 3720
Qy 3721 ctgcacatgcacctggtaacaaagagatagacgttttttcccatggaagggatcggtgggttg 3780
Db 3721 ctgcacatgcacctggtaacaaagagatagacgttttttcccatggaagggatcggtgggttg 3780
Qy 3781 attcgggtcattggaggttttctctacagaaatgtctcggaatcaccctttcaagctcggt 3840
Db 3781 attcgggtcattggaggttttctctacagaaatgtctcggaatcaccctttcaagctcggt 3840

Qy	3841	tcattgtcttagctgctcgtattccctccaaaattaaagcagcagccgagaagatgaaggga	3900	Db	4921		tacagacataattggtcgaatacgaataatccagcaataaaagcagcagcgagtgcaataca	4980
Db	3841	tcattgtcttagctgctcgtattccctccaaaattaaagcagcagccgagaagatgaaggga	3900	Qy	4981		aggagacaaatgccactatttacgagagatgaaaggagcagcttgctgtagggaaaaaagc	5040
Qy	3901	atgttagaagcgttagttgttgaaagatccagaagatgacatctgaacttaaatgaaattc	3960	Db	4981		aggagacaaatgccactatttacgagagatgaaaggagcagcttgctgtagggaaaaaagc	5040
Db	3901	atgttagaagcgttagttgttgaaagatccagaagatgacatctgaacttaaatgaaattc	3960	Qy	5041		ctacaagccagtgggatagctcagaaaaagatgtgagggggaatgaaggggagacaggaac	5100
Qy	3961	cttcgtggcagggaaaaaggttcaacatccatctgacatggaagttcttggggttgatgtg	4020	Db	5041		ctacaagccagtgggatagctcagaaaaagatgtgagggggaatgaaggggagacaggaac	5100
Db	3961	cttcgtggcagggaaaaaggttcaacatccatctgacatggaagttcttggggttgatgtg	4020	Qy	5101		gaacaaaaacaatatgatttccatagactatgaagcaataaagacgtgctagtatcaagc	5160
Qy	4021	gatcaaaagcagcgtgaaggactgttcaaacctctggaattgaaagatttaattctctag	4080	Db	5101		gaacaaaaacaatatgatttccatagactatgaagcaataaagacgtgctagtatcaagc	5160
Db	4021	gatcaaaagcagcgtgaaggactgttcaaacctctggaattgaaagatttaattctctag	4080	Qy	5161		agatttctgaggtctatcaagaaaaaggggatgaataacatgttgccgttacgaataaagg	5220
Qy	4081	agaagagatttcaaaaatttgaaagagaaagtattattatcatcaacaagattctttgatccg	4140	Db	5161		agatttctgaggtctatcaagaaaaaggggatgaataacatgttgccgttacgaataaagg	5220
Db	4081	agaagagatttcaaaaatttgaaagagaaagtattattatcatcaacaagattctttgatccg	4140	Qy	5221		atttcttagaacggatagtttaaagatcatgttggtatcgaccttgaatggttgagagaat	5280
Qy	4141	cgatatttcagtcgtgtggagagttggatccctggttcatgttccaaatcagacgcagagt	4200	Db	5221		atttcttagaacggatagtttaaagatcatgttggtatcgaccttgaatggttgagagaat	5280
Db	4141	cgatatttcagtcgtgtggagagttggatccctggttcatgttccaaatcagacgcagagt	4200	Qy	5281		ctctctgtataaagccaaggactatctcttgagcataaagaggtctctgggttgaagaagt	5340
Qy	4201	ttctcaaacacaggtgtgaacaaaaactgtcagtggaacatcacaaatcagtgcaaaactg	4260	Db	5281		ctctctgtataaagccaaggactatctcttgagcataaagaggtctctgggttgaagaagt	5340
Db	4201	ttctcaaacacaggtgtgaacaaaaactgtcagtggaacatcacaaatcagtgcaaaactg	4260	Qy	5341		ttgaatcgttgacactcttaacactccacaaatctgtcttccctgttgacacgaatttg	5400
Qy	4261	ggagccaaactgtctgtatgaaatttgcttcaagggaatgagagaccgcattctatag	4320	Db	5341		ttgaatcgttgacactcttaacactccacaaatctgtcttccctgttgacacgaatttg	5400
Db	4261	ggagccaaactgtctgtatgaaatttgcttcaagggaatgagagaccgcattctatag	4320	Qy	5401		gaaggatagcagttaggtggatgggtggtctctacaacccctaccctgaatcaactcagt	5460
Qy	4321	aagatctgtgtatgttcagaacacagaacaaactacaatgtcgctcagaagaaacactgatc	4380	Db	5401		gaaggatagcagttaggtggatgggtggtctctacaacccctaccctgaatcaactcagt	5460
Db	4321	aagatctgtgtatgttcagaacacagaacaaactacaatgtcgctcagaagaaacactgatc	4380	Qy	5461		tacacctctggagctatccccagtgctcagtcocatccaaaaattcttctggccaagac	5520
Qy	4381	ttgaaaaaacaaatgaattggaagacactctgtctgttttggctcagccaaagaaatgatacta	4440	Db	5461		tacacctctggagctatccccagtgctcagtcocatccaaaaattcttctggccaagac	5520
Db	4381	ttgaaaaaacaaatgaattggaagacactctgtctgttttggctcagccaaagaaatgatacta	4440	Qy	5521		tttgcaaacactcgaaccaacacactgtatgaattacactaccaactgatttacgttttgaa	5580
Qy	4441	attggcaaacaaactcttccagcagctatgagcagtgctgacactcgacagccacatgtac	4500	Db	5521		tttgcaaacactcgaaccaacacactgtatgaattacactaccaactgatttacgttttgaa	5580
Db	4441	attggcaaacaaactcttccagcagctatgagcagtgctgacactcgacagccacatgtac	4500	Qy	5581		aggtattttgcacaaaagtagaccacaaatgtaatgatgtccaatgagagagagtga	5640
Qy	4501	tagacatagaggattttggaatcaaggtgaaagcccttggttattcttgatgtccatct	4560	Db	5581		aggtattttgcacaaaagtagaccacaaatgtaatgatgtccaatgagagagagtga	5640
Db	4501	tagacatagaggattttggaatcaaggtgaaagcccttggttattcttgatgtccatct	4560	Qy	5641		gaactttgccagtgcttatgtctagtgcaagaactgttttacccggcaccagagagagga	5700
Qy	4561	caccaagagttgacagagtaaaagacaaaaatgtaccacgcaggtttttcagacaaagtg	4620	Db	5641		gaactttgccagtgcttatgtctagtgcaagaactgttttacccggcaccagagagagga	5700
Db	4561	caccaagagttgacagagtaaaagacaaaaatgtaccacgcaggtttttcagacaaagtg	4620	Qy	5701		gcttaacaagtgaactattccggtccctcccgagtccttccctccctgtagccatcccg	5760
Qy	4621	gaagtgttccaagagaattccacaggttcagatcataccatcaacgcctcattgaattaccag	4680	Db	5701		gcttaacaagtgaactattccggtccctcccgagtccttccctccctgtagccatcccg	5760
Db	4621	gaagtgttccaagagaattccacaggttcagatcataccatcaacgcctcattgaattaccag	4680	Qy	5761		tga tagaacctacactctccgttgagaaaaatccctagcaagtgggagcacatcgaaatag	5820
Qy	4681	gaatggagatttcggttctccagcgccttccaaagacacacagcagataccacacata	4740	Db	5761		tga tagaacctacactctccgttgagaaaaatccctagcaagtgggagcacatcgaaatag	5820
Db	4681	gaatggagatttcggttctccagcgccttccaaagacacacagcagataccacacata	4740	Qy	5821		aaaactgtgaaccaataattgaaagccgctcgcgcgggcaagagtgcaactgaataa	5880
Qy	4741	atcaacaagatgagatgaataaaagcatcccatattcaaaaaaacatttttgatctgtca	4800	Db	5821		aaaactgtgaaccaataattgaaagccgctcgcgcgggcaagagtgcaactgaataa	5880
Db	4741	atcaacaagatgagatgaataaaagcatcccatattcaaaaaaacatttttgatctgtca	4800	Qy	5881		ccgagagtgattgaagatgcttactacaatgagaccctgacgagatccacacataa	5940
Qy	4801	actcctctgaagaatgcttcaagacagctggttgatcccgctcagtaaacatcacgagtgct	4860	Db	5881		ccgagagtgattgaagatgcttactacaatgagaccctgacgagatccacacataa	5940
Db	4801	actcctctgaagaatgcttcaagacagctggttgatcccgctcagtaaacatcacgagtgct	4860	Qy	5941		aactcaacattgaaacagtttggaatgactctacgggaacacatggaagaacatcgagc	6000
Qy	4861	gtctaccgagagatagaaactgtgaagcgtggttgatcccgctcagtaaacatcacgagc	4920	Db	5941		aactcaacattgaaacagtttggaatgactctacgggaacacatggaagaacatcgagc	6000
Db	4861	gtctaccgagagatagaaactgtgaagcgtggttgatcccgctcagtaaacatcacgagc	4920	Qy	6001		tccaagaaggtgacatgtccaaggtttggtgtgtttgtatccaacacactcttcttcc	6060
Qy	4921	tacagaacataattggtcgaatccaattccagcaataaaagagacagcgagttgaataca	4980					

QY 301 tgcctactttattgcgtctcttctttaccggttccagctaaacaaattcttcgctattcaat 360
Db 1749 tgcctactttattgcgtctcttctttaccggttccagctaaacaaattcttcgctattcaat 1808
QY 361 gtgtttctcgttttggtagagaaaaatctcgacaaaaatcattttatgcattttat 420
Db 1809 gctgtctcgttttgcgtgagaaaaatctcgacaaaaatcattttatgcattttat 1868
QY 421 ggtgcagattcttagttaatgcgtccctctctaaaccaagtcagattaaaaagagtgcttc 480
Db 1869 ggtgcagattcttagttaatgcgtccctctctaaaccaagtcagattaaaaagagtgcttc 1928
QY 481 gtccatgttgcgtttgtttgtttgagagagttttcggagagttagtgagtggttat 540
Db 1929 gtccatgttgcgtttgtttgtttgagagagttttcggagagttagtgagtggttat 1988
QY 541 ttggggtgaggttagtataagggtttgaaggggagtgattcatcaagtgtgttatgaatt 600
Db 1989 ttggggtgaggttagtataagggtttgaaggggagtgattcatcaagtgtgttatgaatt 2048
QY 601 cgagggtgatccgggggtagatatttccgagttcccttggagaatcaaaactcaacaag 660
Db 2049 cgagggtgatccgggggtagatatttccgagttcccttggagaatcaaaactcaacaag 2108
QY 661 agttcatgggttcttgattccatttacacccaaaaaacctagatcatcaagtctgatgtag 720
Db 2109 agttcatgggttcttgattccatttacacccaaaaaacctagatcatcaagtctgatgtag 2168
QY 721 atgagagagtataaacaggatctaataatgggtttccaggtgtgtaatttgtagacagg 780
Db 2169 atgagagagtataaacaggatctaataatgggtttccaggtgtgtaatttgtagacagg 2228
QY 781 gattctgcacaactggtgtggataataatgggttttcttgatcatggtgcctcatcaggcg 840
Db 2229 gattctgcacaactggtgtggataataatgggttttcttgatcatggtgcctcatcaggcg 2288
QY 841 ttaccaacttaagtatgatcatcaatagcttagcgggatacacatgcacaagcttgagta 900
Db 2289 ttaccaacttaagtatgatcatcaatagcttagcgggatacacatgcacaagcttgagta 2348
QY 901 atagttagagagatcttttggcagagtagtgagtgactctcttttagcaccagttatca 960
Db 2349 atagttagagagatcttttggcagagtagtgagtgactctcttttagcaccagttatca 2408
QY 961 gaaacacacccggttaatgtagacccgtcgaatggaatatttacttcagatgtgggtatgg 1020
Db 2409 gaaacacacccggttaatgtagacccggtcgaatggaatatttacttcagatgtgggtatgg 2468
QY 1021 taaatgctcttccacccagatggcacttctcaagctggctataatgagtttgaaattgg 1080
Db 2469 taaatgctcttccacccagatggcacttctcaagctggctataatgagtttgaaattgg 2528
QY 1081 atgacttgttgaaatcctgatcagatgcccttctccttccacaaagcttgcgtggtgggg 1140
Db 2529 atgacttgttgaaatcctgatcagatgcccttctccttccacaaagcttgcgtggtgggg 2588
QY 1141 atagcttattcaaggttcgtcaatgtgagtgatcaaaatctatttccagtttttttttc 1200
Db 2589 atagcttattcaaggttcgtcaatgtgagtgatcaaaatctatttccagtttttttttc 2648
QY 1201 ccttctcctgttcttcagctacttagatagacatgaattagaatatcttaagaag 1260
Db 2649 ccttctcctgttcttcagctacttagatagacatgaattagaatatcttaagaag 2708
QY 1261 catggttttaacagatggacctccagcgtgtaacaagccttcttaacaattgaattcac 1320
Db 2709 catggttttaacagatggacctccagcgtgtaacaagccttcttaacaattgaattcac 2768
QY 1321 caattagaagagagcagttgggttcagctcgtgaaagtctgtttccaatatgtaccgtcaa 1380
Db 2769 caattagaagagagcagttgggttcagctcgtgaaagtctgtttccaatatgtaccgtcaa 2828
QY 1381 cgccagctcgttccagacacagggtgaacagactggattccttgaacagatagttacaacta 1440

Db 2829 cgccagctcgttccagacaggtgaacagctggtctccttgaacagatagttacaacta 2888
QY 1441 ctggacatgaatcccgagagccgaatactgacaaaaatgacagacattatgacctcgt 1500
Db 2889 ctggacatgaatcccgagagccgaatactgacaaaaatgacagacattatgacctcgt 2948
QY 1501 ctgctgttaatgcgacggaagctactgaacaaaatgatggcagacagaatgtttctgg 1560
Db 2949 ctgctgttaatgcgacggaagctactgaacaaaatgatggcagacagaatgtttctgg 3008
QY 1561 agttcgaccttaacaaactcctcagcagaaaacctccaaaggaaaaggaagtctatgc 1620
Db 3009 agttcgaccttaacaaactcctcagcagaaaacctccaaaggaaaaggaagtctatgc 3068
QY 1621 ccaaggtgcgttggaagcgaacctaataaagaagccacgcaaacctgcgagaacttccca 1680
Db 3069 ccaaggtgcgttggaagcgaacctaataaagaagccacgcaaacctgcgagaacttccca 3128
QY 1681 aagtgtgtcgtgggaagcaaacctaaaagggaagccacgcgaagctgcgaactcaggaaaaag 1740
Db 3129 aagtgtgtcgtgggaagcgaacctaataaagggaagccacgcgaagctgcgaactcaggaaaaag 3188
QY 1741 tgaatctaaagaaaacccgggagtgccaaaaagaaaaatttgaagaatcagcaactaaaa 1800
Db 3189 tgaatctaaagaaaacccgggagtgccaaaaagaaaaatttgaagaatcagcaactaaaa 3248
QY 1801 agccagccaatgttgagatgatgacacaacaaagccctgaagtccactcaaaagtgcga 1860
Db 3249 agccagccaatgttgagatgatgacacaacaaagccctgaagtccactcaaaagtgcga 3308
QY 1861 gaaaagctttgaaatttgcacttggaagatcctggagatgcgaggaaggtgcactctgagt 1920
Db 3309 gaaaagctttgaaatttgcacttggaagatcctggagatgcgaggaaggtgcactctgagt 3368
QY 1921 ctgaaattgtccagaaacagtagtggcgcaaacctgtttcttgagatcagagatgcccattg 1980
Db 3369 ctgaaattgtccagaaacagtagtggcgcaaacctgtttcttgagatcagagatgcccattg 3428
QY 1981 gtggaaactaatgttagttctcctggttagtgcacaataatagacaagaacaaatggattgg 2040
Db 3429 gtggaaactaatgttagttctcctggttagtgcacaataatagacaagaacaaatggattgg 3488
QY 2041 gggctatgaacccagccacttgaagtgcgaatgggaacccagccagataaactatctcac 2100
Db 3489 gggctatgaacccagccacttgaagtgcgaatgggaacccagccagataaactatctcac 3548
QY 2101 gaggaaaactggccagagaccaaaccctgattttattgactagaaaaccagaaatggccagt 2160
Db 3549 gaggaaaactggccagagaccaaaccctgattttattgactagaaaaccagaaatggccagt 3608
QY 2161 tcccagtggaacccagaaacacccagttcccgaatggaaaaccaacagcttggcttcaga 2220
Db 3609 tcccagtggaacccagaaacacccagttcccgaatggaaaaccaacagcttggcttcaga 3668
QY 2221 tgaaaaaccaactatttggctttccatttggtaaccagcaacctcgcatgaccataagaa 2280
Db 3669 tgaaaaaccaactatttggctttccatttggtaaccagcaacctcgcatgaccataagaa 3728
QY 2281 accagcagccttgccttggccatgggttaatacaaacacctatgtatctctgataggaaactccac 2340
Db 3729 accagcagccttgccttggccatgggttaatacaaacacctatgtatctctgataggaaactccac 3788
QY 2341 ggctgatttagtaagtggaaacacagcaactaggaggtcccgaaggaaaacagcggccta 2400
Db 3789 ggctgatttagtaagtggaaacacagcaactaggaggtcccgaaggaaaacagcggccta 3848
QY 2401 tatttttgaatcaccaagcttgtttaccctgtggaatcagctatatgtgatccactcac 2460
Db 3849 tatttttgaatcaccaagcttgtttaccctgtggaatcagctatatgtgatccactcac 3908
QY 2461 acatgcatcaacttgttatgtcaacccggagggaacaaactggactactgataaaaaacc 2520

Qy	4280	tgaatttgccttccaagggaaatgagagaccgcacatctatatgaaaggactctggtgatgttca	4339
Db	6129	tgaatttgccttccaagggaaatgagagaccgcacatctatatgaaaggactctggtgatgttca	6188
Qy	4340	gaacaagaanaactacaaatgtcgctcagaagaanaacctgatcttgaanaaacaatgaattg	4399
Db	6189	gaacaagaanaactacaaatgtcgctcagaagaanaacctgatcttgaanaaacaatgaattg	6248
Qy	4400	gaagactctgtctgttttggctcagccaagaataatgatactacttggcaacaactccttc	4459
Db	6249	gaagactctgtctgttttggctcagccaagaataatgatactacttggcaacaactccttc	6308
Qy	4460	cagcagctatgagcagctgcgactcgacgcagccacatgtactagacatagaggattttgg	4519
Db	6309	cagcagctatgagcagctgcgactcgacgcagccacatgtactagacatagaggattttgg	6368
Qy	4520	aatgcgaagtgaaagccttggttatcttcttgatgtccatctcaccagagagttgaacaggt	4579
Db	6369	aatgcgaagtgaaagccttggttatcttcttgatgtccatctcaccagagagttgaacaggt	6428
Qy	4580	aasgaacaaaaatgtaccacgaggtttttcagacaaggtggaagtggtccaagagaatt	4639
Db	6429	aasgaacaaaaatgtaccacgaggtttttcagacaaggtggaagtggtccaagagaatt	6488
Qy	4640	cacagctcagatcatcaccatcaacgcctcatgaattaccaggaattggattgcgcggttc	4699
Db	6489	cacagctcagatcatcaccatcaacgcctcatgaattaccaggaattggattgcgcggttc	6548
Qy	4700	ctcaagcgccgtccaagaacacacagagagatataccaaacataatacaacaagatgagatgaa	4759
Db	6549	ctcaagcgccgtccaagaacacacagagagatataccaaacataatacaacaagatgagatgaa	6608
Qy	4760	taaagatccattttcaaaaaaacatttttgatctgctcaactcctctgaagaatgcct	4819
Db	6609	taaagatcccatcttcaaaaaaacatttttgatctgctcaactcctctgaagaatgcct	6668
Qy	4820	tacaagacgtccagtaccaaacagaaacatcacgagtggtctctaccgagagatagaac	4879
Db	6669	tacaagacgtccagtaccaaacagaaacatcacgagtggtctctaccgagagatagaac	6728
Qy	4980	tgctgaagacgtggttgatccgctcagtaacaattcaagcttacagaacaatatgtgtcga	4939
Db	6729	tgctgaagacgtggttgatccgctcagtaacaattcaagcttacagaacaatatgtgtcga	6788
Qy	4940	atcaaatccagcaataaagagacagacgagcagttgaaatacaagagacaaaatgccactat	4999
Db	6789	atcaaatccagcaataaagagacagacgagcagttgaaatacaagagacaaaatgccactat	6848
Qy	5000	tttacgagagatgaaggagcagcttgctgatgggaaaaagcctcaagccagtgggatag	5059
Db	6849	tttacgagagatgaaggagcagcttgctgatgggaaaaagcctcaagccagtgggatag	6908
Qy	5060	tctcagaaaaagatgtggaggggaatgaaggagacaggaacgaaaaaacaatatggga	5119
Db	6909	tctcagaaaaagatgtggaggggaatgaaggagacaggaacgaaaaaacaatatggga	6968
Qy	5120	ttccatagactatgaagcaataaagacgtgctagtatcgccagatcttctgaggtatcaa	5179
Db	6969	ttccatagactatgaagcaataaagacgtgctagtatcgccagatcttctgaggtatcaa	7028
Qy	5180	ggaaagaggatgaataaacatgttgcccgtagcaattaaagatttccctagacagcgatag	5239
Db	7029	ggaaagaggatgaataaacatgttgcccgtagcaattaaagatttccctagacagcgatag	7088
Qy	5240	taa 5242	
Db	7089	tga 7091	
RESULT 3			
AAS96686			
ID AAS96686 standard; DNA; 12785 BP.			
XX			

AAS96686;
 26-FEB-2002 (first entry)
 Arabidopsis DNA for Demeter protein.
 Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ds;
 DNA demethylation; transgenic plant; transcription modulation;
 flowering time; endosperm development; MEDEA.
 Arabidopsis Thaliana.
 W0200180626-A1.
 01-NOV-2001.
 23-APR-2001; 2001WO-US13059.
 21-APR-2000; 2000US-0553690.
 (REGC) UNIV CALIFORNIA.
 Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
 WPI; 2002-055307/07.
 P-PSDB; AAU72736.
 New polynucleotide that control plant development comprising a sequence
 having a specific homology to DEMETER domains A,B or C
 Example 1; Page 46-54; 109pp; English.
 The invention relates to an isolated polynucleotide sequence or their
 complement encoding a polypeptide having a sequence at least 40%
 identical to DMT (DEMETER, previously known as ATRPOS (ATR)) Domain A,
 or C or their combinations. Also included are an expression
 cassette comprising the polynucleotide or comprising a heterologous
 polynucleotide under the control of a promoter at least 70% identical to
 DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
 region of DMT, a host cell comprising an exogenous polynucleotide
 encoding a DMT-like protein and a transgenic plant comprising a
 polynucleotide encoding a DMT-like protein. The expression cassette is
 useful for modulating transcription. The method comprises introducing
 the cassette into a host cell preferably Agrobacterium by sexual
 cross, and selecting a host cell with modulated transcription, where
 the protein is capable of exhibiting at least one of the following
 biological activities, which include enhanced expression of the
 protein in a plant results in a delay in flowering time, introduction of
 the protein into a cell results in modulation of methylation of
 chromosomal DNA in the cell, reduction of expression of the protein in a
 plant results in enhanced endosperm development and expressing of the
 protein in an Arabidopsis leaf results in expression of the MEDEA gene.
 The polynucleotide is useful for detecting a nucleic acid in a sample.
 DEMETER is related to 5-methylcytosine glycosylases and regulates
 transcription of target genes by demethylation. The present
 sequence encodes a DMT protein.
 Sequence 12785 BP; 4320 A; 2315 C; 2496 G; 3654 T; 0 other

	Query Match	69.7%;	Score 4788.6;	DB 24;	Length 12785;
	Best Local Similarity	92.6%;	Pred. No. 0;		
	Matches 5228;	Conservative 0;	Mismatches 14;	Indels 401;	Gaps 3;
<hr/>					
Qy	1	gtttctcggcgattgactgcgctgagaatcagaaaagttagatcggtagctttagctcc	60		
Db	3425	gtttccggcattgactgcgctgagaatcagaaagcttagatcggtagctttagctcc	3484		
<hr/>					
Qy	61	atttcctgtttattcacatatattctctttttctctcccccttttttatctgggaatt	120		
Db	3485	atttcctgtttattcacatatattctctttttctctcccccttttttatctgggaatt	3544		
<hr/>					
Qy	121	tgttcgtgtaaattttcacgctgttacattttccgatcacgagaagaatactagggtttt	180		

QY	2341	ggcctgcattagtaagttggaacccagacaactagaggtgccccaaggaaacaagcggccta	2400
DB	5765	ggcctgcattagtaagttggaacccagacaactagaggtgccccaaggaaacaagcggccta	5824
QY	2401	tatttttgatcaccagactgtttacctgctggaatcagctatatggatcacctacag	2460
DB	5825	tatttttgatcaccagactgtttacctgctggaatcagctatatggatcacctacag	5884
QY	2461	acaatgcatacaacttgttatgtcaaccggaggggcaacaacatggactactgataaaaaacc	2520
DB	5885	acaatgcatacaacttgttatgtcaaccggaggggcaacaacatggactactgataaaaaacc	5944
QY	2521	agaaacctggatcattataagagccagcagccttgcgtacctttgattgaccagcaac	2580
DB	5945	agaaacctggatcattataagagggccagcagccttgcgtacctttgattgaccagcaac	6004
QY	2581	ctgaaactccaaaaggttttactcaacttgaatcagatggtagctacagcatgcatcgc	2640
DB	6005	ctgaaactccaaaaggttttactcaacttgaatcagatggtagctacagcatgcatcgc	6064
QY	2641	ctggggttcgaactcatctcgaatcaaaagttcctacaacaatatctcatgttggaatctg	2700
DB	6065	ctggggttcgaactcatctcgaatcaaaagttcctacaacaatatctcatgttggaatctg	6124
QY	2701	ttccagagattttgaatgggactacaggtacatgccagagaaagcagggtccttgcatacgg	2760
DB	6125	ttccagagattttgaatgggactacaggtacatgccagagaaagcagggtccttgcatacgg	6184
QY	2761	attctttacagcaagatattccatcaaggaaataagtagcatatactttctatgagatcca	2820
DB	6185	attctttacagcaagatattccatcaaggaaataagtagcatatactttctatgagatcca	6244
QY	2821	atggttaatgggtgcagaagaagcgttaactcaaaaactcttctgcgaactccaattatgg	2880
DB	6245	atggttaatgggtgcagaagaagcgttaactcaaaaactcttctgcgaactccaattatgg	6304
QY	2881	ctaaacttgaggaagccaggggctcgaaagagacagtagtcatcgtgcgaatgggacagcgg	2940
DB	6305	ctaaacttgaggaagccaggggctcgaaagagacagtagtcatcgtgcgaatgggacagcgg	6364
QY	2941	aaaagcatgatctaaacttagtctcaacagatgtgtccaatcacagatgtggagagacata	3000
DB	6365	aaaagcatgatctaaacttagtctcaacagatgtgtccaatcacagatgtggagagacata	6424
QY	3001	acagcagcacgctgtgtggaataattagatgtctgcaaaaagaaacagaaatccagaaagtag	3060
DB	6425	acagcagcacgctgtgtggaataattagatgtctgcaaaaagaaacagaaatccagaaagtag	6484
QY	3061	tccaagaaaatttgcagtgcacatgccacctggaggttatagaatcgagatgattcccaactg	3120
DB	6485	tccaagaaaatttgcagtgcacatgccacctggaggttatagaatcgagatgattcccaactg	6544
QY	3121	atggggcaagaaaaggttaaaaatactgccagcatcagtaaaaggtgtcatctaaaggaact	3180
DB	6545	atggggcaagaaaaggttaaaaatactgccagcatcagtaaaaggtgtcatctaaaggaact	6604
QY	3181	cgctccagttaaaagacagcagaaaaggagaaatgtattgtcccaaaaacgcctgcaa	3240
DB	6605	cgctccagttaaaagacagcagaaaaggagaaatgtattgtcccaaaaacgcctgcaa	6664
QY	3241	aaaaggttcagcaggttagaaaaataatcagtaacctccgcctgtctcatgctccagagatcc	3300
DB	6665	aaaaggttcagcaggttagaaaaataatcagtaacctccgcctgtctcatgctccagagatcc	6724
QY	3301	agctttggcaacttactctctccaaagacacctttatcaagaagcaagcctaaaggaag	3360
DB	6725	agctttggcaacttactctctccaaagacacctttatcaagaagcaagcctaaaggaag	6784
QY	3361	ggagaaggtccatacaagattcaggaaaagcaag-----	3394
DB	6785	ggagaaggtccatacaagattcaggaaaagcaaggttaactaatgtattctacaatctc	6844

Db 7925 agaagagaaagtattatcatcacaaagattctttgatccggcgatatttcagtcgtgtgg 7984
Qy 4160 gagagttggtacctgttcatttccaaatcagacgacgaggtttctctacaaccaggtgtga 4219
Db 7985 gagagttggtacctgttcatttccaaatcagacgacgaggtttctctacaaccaggtgtga 8044
Qy 4220 aacaaaactgtcagtggaacatcacatcagtcgaaactgggagcccaaaacttgtctga 4279
Db 8045 aacaaaactgtcagtggaacatcacatcagtcgaaactgggagcccaaaacttgtctga 8104
Qy 4280 tgaattttgtctcaagggaatgagacccgcatctatatgaaggatctgtgtatgttca 4339
Db 8105 tgaattttgtctcaagggaatgagacccgcatctatatgaaggatctgtgtatgttca 8164
Qy 4340 gaaacagaactacaaatgcgtcgaagaaacctgatcttgaaaaaacaatgaattg 4399
Db 8165 gaaacagaactacaaatgcgtcgaagaaacctgatcttgaaaaaacaatgaattg 8224
Qy 4400 gaaagactctgtctgtttgttcgacgaagaatgatactaatggcaacaactcttc 4459
Db 8225 gaaagactctgtctgtttgttcgacgaagaatgatactaatggcaacaactcttc 8284
Qy 4460 cagcagctatgagcagtggtcgactcgacagccacatgtactagacatagaggattttgg 4519
Db 8285 cagcagctatgagcagtggtcgactcgacagccacatgtactagacatagaggattttgg 8344
Qy 4520 aatcgaagtggaagccctgtgttcttctgtgatgtccatctccaaagagtgacagagt 4579
Db 8345 aatcgaagtggaagccctgtgttcttctgtgatgtccatctccaaagagtgacagagt 8404
Qy 4580 aagaacaaaatgtaccacgcaggtttttcagacaaggtggaagtgttccaaagaaatt 4639
Db 8405 aagaacaaaatgtaccacgcaggtttttcagacaaggtggaagtgttccaaagaaatt 8464
Qy 4640 cacaggtcagatcattacatcaacgcctcatgaattaccaggaatggattgtccggttc 4699
Db 8465 cacaggtcagatcattacatcaacgcctcatgaattaccaggaatggattgtccggttc 8524
Qy 4700 ctcaagcgccgtccaagaacaccaggacgataccacaataatcaacaagatgagatgaa 4759
Db 8525 ctcaagcgccgtccaagaacaccaggacgataccacaataatcaacaagatgagatgaa 8584
Qy 4760 taagatcccatatttacaacaaacatttttggtatctgtcctcaactctctgaagaatgcct 4819
Db 8585 taagatcccatatttacaacaaacatttttggtatctgtcctcaactctctgaagaatgcct 8644
Qy 4820 taacagcagtcagttaccacaaacagacatcacggatggctgtctacggagagatagaac 4879
Db 8645 taacagcagtcagttaccacaaacagacatcacggatggctgtctacggagagatagaac 8704
Qy 4880 tgcgtgaagcgtggtgatccgctcagtaacaaattcaagcttcagaaacatattggtcga 4939
Db 8705 tgcgtgaagcgtggtgatccgctcagtaacaaattcaagcttcagaaacatattggtcga 8764
Qy 4940 atcaaatccagcaataaagagcagcggcagttgaaatacaaggagacaaatgcacctat 4999
Db 8765 atcaaatccagcaataaagagcagcggcagttgaaatacaaggagacaaatgcacctat 8824
Qy 5000 ttacgagagatgaaaggacgttgtgtgatgggaaagcctcacaagcagtggtgatag 5059
Db 8825 ttacgagagatgaaaggacgttgtgtgatgggaaagcctcacaagcagtggtgatag 8884
Qy 5060 tctcagaaaaagatgtgagggggaatgaaggagagacgagaaacgaaacaaacatatgga 5119
Db 8885 tctcagaaaaagatgtgagggggaatgaaggagagacgagaaacgaaacaaacatatgga 8944
Qy 5120 ttccatagactatgaagcaataagacgtgttagtatcagcagagattttctgagggtatcaa 5179
Db 8945 ttccatagactatgaagcaataagacgtgttagtatcagcagagattttctgagggtatcaa 9004
Qy 5180 ggaagaggagatgaataacatgttggcgcgtacgaattaaagagatttcttagaacggtatgt 5239
|||||

Db 9005 ggaagagggatgaataacatgttggcgcgtacgaattaagtaaatctactaatttcagt 9064
Qy 5240 taa 5242
Db 9065 tga 9067
RESULT 4
AAS96690
ID AAS96690 standard; CDNA; 1478 BP.
XX
AC AAS96690;
XX
DT 26-FEB-2002 (first entry)
XX
DE Arabidopsis DMT 5' untranslated region.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA; 5' untranslated region.
XX
OS Arabidopsis Thaliana.
XX
PN W0200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamura JK, Tatarinova TV;
XX
XX WPI; 2002-055307/07.
PT New polynucleotide that control plant development comprising a sequence
PT having a specific homology to DEMETER domains A,B or C -
PS Claim 11; Page 62; 109pp; English.
XX The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence represents the 5' untranslated region of the Arabidopsis
CC DMT gene.
XX
SQ Sequence 1478 BP; 392 A; 261 C; 322 G; 503 T; 0 other;

Query Match 21.5%; Score 1478; DB 24; Length 1478;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttctccggcaattgactcgcctgagaatcagaagacttagatcggtgagcttttagctcc 60
Db 1 gttctccggcaattgactcgcctgagaatcagaagacttagatcggtgagcttttagctcc 60
Qy 61 attttctggtttattacatatattttctctttttttttctctctctctcttttttttttgaatt 120
Db 61 attttctggtttattacatatattttctctttttttttttctctctctctcttttttttttgaatt 120
Qy 121 tgttctgctaaattttccagctgttacattttccgataccagagagaatacaactgggtttt 180
Db 121 tgttctgctaaattttccagctgttacattttccgataccagagagaatacaactgggtttt 180
Qy 181 tatgttaatacaatcatgttctctgtttttctgatacaataatcctcagctattaaacctgat 240
Db 181 tatgttaatacaatcatgttctctgtttttctgatacaataatcctcagctattaaacctgat 240
Qy 241 tttgattctgcgtaataaaacctctgatttggtttttatctttcatttccactttccccaataaact 300
Db 241 tttgattctgcgtaataaaacctctgatttggtttttatctttcatttccactttccccaataaact 300
Qy 301 tgcattacttattcgcctctcttttaccgtttccagctaaataattttcgcatttcaat 360
Db 301 tgcattacttattcgcctctcttttaccgtttccagctaaataattttcgcatttcaat 360
Qy 361 gtgtttctcgttttggatgagaaaaatattctgacaaaaaatcatttattgcaattttat 420
Db 361 gtgtttctcgttttggatgagaaaaatattctgacaaaaaatcatttattgcaattttat 420
Qy 421 ggtgcagattcttagttaatgctcctctctcttaaccagtcagattaaaaaggagtgctc 480
Db 421 ggtgcagattcttagttaatgctcctctctcttaaccagtcagattaaaaaggagtgctc 480
Qy 481 gtcattgttctgttttgggtgttgtagagagatttttcggagaggttagtgagtggtat 540
Db 481 gtcattgttctgttttgggtgttgtagagagatttttcggagaggttagtgagtggtat 540
Qy 541 ttgggttgaggtgagtaagtttgaagggtgaggtgattcattcaatggtgttatgaatt 600
Db 541 ttgggttgaggtgagtaagtttgaagggtgaggtgattcattcaatggtgttatgaatt 600
Qy 601 cgagggctgataccgggggatagatttttcgaggttctcttggagaaatcaaaactcaacaag 660
Db 601 cgagggctgataccgggggatagatttttcgaggttctcttggagaaatcaaaactcaacaag 660
Qy 661 agttcattggttctgtgattccatttaccaccccaaaaacctagatcaagtcctgatgtag 720
Db 661 agttcattggttctgtgattccatttaccaccccaaaaacctagatcaagtcctgatgtag 720
Qy 721 atgagagagtataaaccagatctaattgggtttccaggtggtgaaattttgtagacaggg 780
Db 721 atgagagagtataaaccagatctaattgggtttccaggtggtgaaattttgtagacaggg 780
Qy 781 gattctgcaacactggtgtggatcataatgggggttttttgatcatggtgctcactcaggggc 840
Db 781 gattctgcaacactggtgtggatcataatgggggttttttgatcatggtgctcactcaggggc 840
Qy 841 ttaccacttaagtatgatcataatagcttagcggatcacatgcacaagccttgagta 900
Db 841 ttaccacttaagtatgatcataatagcttagcggatcacatgcacaagccttgagta 900
Qy 901 atagtggagagatcttttgggcaggagtgaggtgacttctctcttttagcaccagttatca 960
Db 901 atagtggagagatcttttgggcaggagtgaggtgacttctctcttttagcaccagttatca 960
Qy 961 gaaacacacccggtaattgtagagccgttcaatggaaatttttacttcagatgtgggtatgg 1020
Db 961 gaaacacacccggtaattgtagagccgttcaatggaaatttttacttcagatgtgggtatgg 1020
Qy 1021 taaatggtctttcaccacagtgagcacttctcgaagctggtcactaaatgagttgaaattgg 1080
Db 1021 taaatggtctttcaccacagtgagcacttctcgaagctggtcactaaatgagttgaaattgg 1080

Qy 1081 atgacttgttgaaactcctgatacagatgcccttctctcttcacaaagcttctgctgagtggtggg 1140
Db 1081 atgacttgttgaaactcctgatacagatgcccttctctcttcacaaagcttctgctgagtggtggg 1140
Qy 1141 atagcttattcaagggttcgtaactgtgagtgatcaaaatctatttttcttttttttc 1200
Db 1141 atagcttattcaagggttcgtaactgtgagtgatcaaaatctatttttcttttttttc 1200
Qy 1201 cctttctcctgttctcagtaactacttagatagaacatgaataatcttaagaagt 1260
Db 1201 cctttctcctgttctcagtaactacttagatagaacatgaataatcttaagaagt 1260
Qy 1261 catggttttgaacagatggacctccagcgtgttaacagcctctttacaatttgaattcac 1320
Db 1261 catggttttgaacagatggacctccagcgtgttaacagcctctttacaatttgaattcac 1320
Qy 1321 caattagaagagaagcagttgggtcagctctgtgaaagtctgttcaatgtaccgtcaa 1380
Db 1321 caattagaagagaagcagttgggtcagctctgtgaaagtctgttcaatgtaccgtcaa 1380
Qy 1381 cgcccgatctgttcagaaacaggtgaaagactgattccttgaaacagatagttacaacta 1440
Db 1381 cgcccgatctgttcagaaacaggtgaaagactgattccttgaaacagatagttacaacta 1440
Qy 1441 ctggacatgaaatcccgagagccgaaatctgacaaaagt 1478
Db 1441 ctggacatgaaatcccgagagccgaaatctgacaaaagt 1478
RESULT 5
AAS96707
ID AAS96707 standard; cDNA; 758 BP.
AC AAS96707;
XX
DT 26-FEB-2002 (first entry)
XX
DE Tomato DMT related cDNA sequence #1.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
KW DNA demethylation; transgenic plant; transcription modulation;
XX flowering time; endosperm development; MEDEA.
OS Lycopersicon esculentum.
XX
PN WO200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okumuro JK, Tatarinova TV;
XX
DR WPI; 2002-055307/07.
DR P-PSDB; AAU72757.
XX
PT New polynucleotide that control plant development comprising a sequence
PS having a specific homology to DEMETER domains A,B or C .
XX
PS Disclosure; Page 92-93; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated

Db	742	tattcagagtttccctgaattctattggtgacaagcatgggggcatcgattcttgatggct	801
Qy	5273	gagagaattctctctgtataaagccaaagactatctcttggagcataagaggtctgggttt	5332
Db	802	gagagatgttccacctgatcaagcaaaagaattcttgcagcataaggggattgggatt	861
Qy	5333	gaaaagtgttgatcggtcgactcttaacactccacaattctgtcttccctgttgacac	5392
Db	862	gaaaagtggaggtgtcagactcttaacactcaccatcttgccttccogtgggac	921
Qy	5393	gaagtgtgaagatagcagttaggatggatgggtgcctctcaaacccctacctgaac	5452
Db	922	aaatgttggcgtatagcagtgaaatctggatcgggtgcctccagcactgccagagtc	981
Qy	5453	acttcagttacacctcctggagctatccagctgcctgagtcctccaaaattctctt	5512
Db	982	actacagttgcattctctagaattgtaccagctgtggagtcctacacaaaatactct	1041
Qy	5513	gccaaagactttgcaactcgatacgaacactgtatgaattacaactaccacactgatt	5572
Db	1042	gccccgctgcgaagctagacaaagaacattgtatgagctgcattaccagctgattac	1101
Qy	5573	gtttgaaaagtattttgcacaaagtagacc	5605
Db	1102	atttggaaaggtctctgtactaaaagcaagcc	1134
RESULT 10			
AAS96704			
ID	AAS96704 standard; cDNA; 616 BP.		
AC	AAS96704;		
XX			
DT	26-FEB-2002 (first entry)		
XX			
DE	Soybean DMT3 DNA.		
XX			
KW	Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;		
KW	DNA demethylation; transgenic plant; transcription modulation;		
KW	flowering time; endosperm development; MEDEA.		
OS	Glycine max.		
XX			
PN	WO200180626-A1.		
XX			
PD	01-NOV-2001.		
XX			
PF	23-APR-2001; 2001WO-US13059.		
XX			
PR	21-APR-2000; 2000US-0553690.		
XX			
PA	(REGC) UNIV CALIFORNIA.		
XX			
PI	Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;		
XX			
DR	WPI; 2002-055307/07.		
DR	P-PSDB; AAU72754.		
XX			
PT	New polynucleotide that control plant development comprising a sequence		
PT	having a specific homology to DEMETER domains A,B or C		
XX			
PS	Disclosure; Page 90-91; 109pp; English.		
XX			
CC	The invention relates to an isolated polynucleotide sequence or their		
CC	complement encoding a polypeptide having a sequence at least 40%		
CC	identical to DMT (DEMETER, previously known as AtRPO5 (ATR)) Domain A, B		
CC	or C or their combinations. Also included are an expression		
CC	cassette comprising the polynucleotide or comprising a heterologous		
CC	polynucleotide under the control of a promoter at least 70% identical to		
CC	DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated		
CC	region of DMT, a host cell comprising an exogenous polynucleotide		
CC	encoding a DMT-like protein and a transgenic plant comprising a		
CC	polynucleotide encoding a DMT-like protein. The expression cassette is		

CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expression of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence encodes a DMT-like protein.
XX
SQ Sequence 616 BP; 211 A; 137 C; 131 G; 135 T; 2 other;

Query Match 3.7%; Score 253; DB 24; Length 616;
Best Local Similarity 67.2%; Pred. NO. 6.2e-54;
Matches 385; Conservative 2; Mismatches 182; Indels 4; Gaps 2;
QY 5802 ggagcaccatcgatagagaaactgtgaaccaataattgaagagccgctcgccggg 5861
Db 47 gaagcaaatctgaatacaatcgaatcgcaacacctattattgaagagccgcaactccagag 106
QY 5862 caagagtcgactgaataaccgagagtgatattgaagatgcttactacaatgagaccct 5921
Db 107 ccagaatgctccagatattccgaaatgatagaggatac---cttcaatgaggaaatca 163
QY 5922 gacgagatcccaataataaaactcaacattgaacagtttggaatgactctacgggaacac 5981
Db 164 tgtgaattcccaacatcaaatcagatagagaagtgacttgaacttacaacaaactat 223
QY 5982 atggaagaagaactgagctccaagaagtgcactgtccaagccttggtgcttggat 6041
Db 224 atgcaagaagaacatggaacttcaagaaggtgaaatgtcaaaaggccttggtgcttcatat 283
QY 6042 ccaacaactactctatttccaaactccaaactaaagaacattagcgcgtctcaggacagag 6101
Db 284 ccaggtgctgcattctactacccaagctgaagaatgtgagcgggttgcaaacagag 343
QY 6102 caacaagtgacagctcccaattcaacatgctctccttgatggtatgataaaagagaa 6161
Db 344 cattatgtttatgaactccctgattcacatccctctctgaatgggtggaacaagcgagaa 403
QY 6162 ccagatgatcccaagtccttctcttctatgctatggaacacagtggaacaagcaattcg 6221
Db 404 cctgatgatccaggcaatacctctctgattggttccctccaggggagacagcagattct 463
QY 6222 gcacaaccgctgaacagaagtggtggaggaaagcgtctggcaaaatgtgctttgacgag 6281
Db 464 atacagccaccagaagaacaaatgcagctctcaggaaatgtggcggctctgtaagtgaat 523
QY 6282 acttttttgatgttaacagcttgaggagaacaa--actcaacagacagctcgaggaactct 6340
Db 524 gaatttttctatgcacagtttccgtgaagcaaggtttcacagatagttcgaggacact 583
QY 6341 tctgataccttgcgactgcccattgagaggaag 6373
Db 584 cctgatacaatgctgaacagctwtgagggag 616

RESULT 11
AAS96696
ID AAS96696 standard; cDNA; 798 BP.
XX
AC AAS96696;
XX
DT 26-FEB-2002 (first entry)
XX
DE Corn DMT1 DNA.
XX

KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
KW DNA demethylation; transgenic plant; transcription modulation;
XX flowering time; endosperm development; MEDEA.
OS Zea mays.
XX
PW WO200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (RECC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
DR WPI; 2002-055307/07.
DR P-PSDB; AAU72745.
XX
PT New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
PS Disclosure; Page 84-85; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expression of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence encodes a DMT-like protein.
XX
SQ Sequence 798 BP; 239 A; 159 C; 188 G; 212 T; 0 other;

Query Match 3.4%; Score 233; DB 24; Length 798;
Best Local Similarity 66.3%; Pred. NO. 8.6e-49;
Matches 335; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
QY 6159 gaaccagatgatccaaagtccttattcttagctatatatggaacaccaggtgaacacggaact 6218
Db 1 gaaccagatgatccttgcctcatatcttccattatgagccccaggtgaaactgcacaa 60
QY 6219 tcggcacaccgcctgaacagaagtggtggaggaaagcgtctggcaaaatgtgctttgac 6278
Db 61 tcatcgtatgcccccaagacattctgtattcaggggagcgggtgagactatgtggaagt 120
QY 6279 gagactttcttgagttaacagctctgagggaagcaaaactcacacagcttcgaggaact 6338
Db 121 tcaacatgcttttagttgcaacaataacagaaaatgcaggctcagaagaatcagaggaaca 180
QY 6339 cttctgataccttgcgactgcgaatgagaggaagtcttcgctcaacggaacatatttc 6398
Db 181 cttttgataccatgcgcgaacagcaatgagaggaagcttcccaacttaattggagcgtatttt 240

CC	protein in a plant results in a delay in flowering time, introduction of the protein into a cell results in modulation of methylation of chromosomal DNA in the cell, reduction of expression of the protein in a plant results in enhanced endosperm development and expressing of the protein in an Arabidopsis leaf results in expression of the MEDEA gene. The polynucleotide is useful for detecting a nucleic acid in a sample. DEMETER is related to 5-methylcytosine glycosylases and regulates transcription of target genes by demethylation. The present sequence encodes a DMT-like protein.
XX	Sequence 517 BP; 150 A; 110 C; 125 G; 132 T; 0 other;
SQ	Query Match Best Local Similarity 3.3%; Score 228.2; DB 24; Length 517; Matches 343; Conservative 66.3%; Pred. No. 1.1e-47; Mismatches 173; Indels 1; Gaps 1;
QY	6111 tacgagctccagattcaca-tcgtctcttctggtgattgataaaagagaacacgatga 6169
DB	1 tatgaactgccagattcacacgcctctcttggaagattcgaaacagagagaacacgatga 60
QY	6170 tccaagtcccttattctatgatatgacacacaggtgaaacagcgaaattcgcgcaaac 6229
DB	61 tcctgtccatattcttcttccataggaccaggtgaaactgcacatcgcgatgc 120
QY	6230 gcctgaacagaagtgaggagaaagcgtctgcgcaaaatgtctttgacgagactgttc 6289
DB	121 ccccaagacattctgtgattccaggggagacggttagactatgtggaagtccaacatgctt 180
QY	6290 tgaggttaacagctgagggaagcaaaactcacagacagcttcgagggaactcttctgatacc 6349
DB	181 tagttgcaacaataatacagagaatgcaggctcagaagaatcagaggaacacttttgatacc 240
QY	6350 ttgtcgaactgcctatgagagaaagtcttcgcgtcaacggagacatatattccaagtcaacga 6409
DB	241 atgcgcgaacgaatgagaggaagcttccacacttaatggagcgtattttcaagttcaatga 300
QY	6410 gttatttgcagaccagcagtcagctcacaacccatcgatgttctctagagattgatatg 6469
DB	301 ggtatttgcgtgacctattgctcaagtcacaaatccaattgattgctccacgaagtgtgattg 360
QY	6470 ggatctcccaagaaggactgtttacttcggaacatcagtaacatcaatattcagaaggtct 6529
DB	361 ggacctcccaagacgaaactgtttactttggaacctcagttctctacaatattcagaaggttt 420
QY	6530 tccaacggagcagatatacagttctcttcttttggaaaggattcgtatgtcgtggattcga 6589
DB	421 aacgactgaagagatacaacagatgcttttggagaggatttgttgcgtgaggggctttga 480
QY	6590 acagaagacaagagcagcgcgtccatttaattggcaagg 6626
DB	481 taggacagtcgaggcaccacgaaggccctttatgcaagg 517
RESULT 13	
AAS96711	
ID	AAS96711 standard; cDNA; 583 BP.
XX	
AC	AAS96711;
DT	26-FEB-2002 (first entry)
XX	
DE	Soybean DMT related cDNA sequence.
XX	Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
KW	DNA demethylation; transgenic plant; transcription modulation;
KW	flowering time; endosperm development; MEDEA.
XX	Glycine max.
OS	
XX	WO200180626-A1.
PN	
XX	01-NOV-2001.
PD	
XX	
XX	
PR	23-APR-2001; 2001WO-US13059.
XX	
XX	21-APR-2000; 2000US-0553690.
PA	(REGC) UNIV CALIFORNIA.
XX	
PI	Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX	
DR	WPI; 2002-055307/07.
DR	P-PSDB; AAU72759.
XX	
PT	New polynucleotide that control plant development comprising a sequence having a specific homology to DEMETER domains A,B or C
XX	
PS	Disclosure; Page 94; 109pp; English.
XX	
CC	The invention relates to an isolated polynucleotide sequence or their complement encoding a polypeptide having a sequence at least 40% identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B or C or their combinations. Also included are an expression cassette comprising the polynucleotide or comprising a heterologous polynucleotide under the control of a promoter at least 70% identical to DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated region of DMT, a host cell comprising an exogenous polynucleotide encoding a DMT-like protein and a transgenic plant comprising a polynucleotide encoding a DMT-like protein. The expression cassette is useful for modulating transcription. The method comprises introducing the cassette into a host cell preferably Agrobacterium by sexual cross, and selecting a host cell with modulated transcription, where the protein is capable of exhibiting at least one of the following biological activities, which include enhanced expression of the
CC	protein in a plant results in a delay in flowering time, introduction of the protein into a cell results in modulation of methylation of chromosomal DNA in the cell, reduction of expression of the protein in a plant results in enhanced endosperm development and expressing of the protein in an Arabidopsis leaf results in expression of the MEDEA gene. The polynucleotide is useful for detecting a nucleic acid in a sample. DEMETER is related to 5-methylcytosine glycosylases and regulates transcription of target genes by demethylation. The present sequence encodes a DMT-like protein.
XX	Sequence 517 BP; 150 A; 110 C; 125 G; 132 T; 0 other;
SQ	Query Match Best Local Similarity 3.3%; Score 228.2; DB 24; Length 517; Matches 343; Conservative 66.3%; Pred. No. 1.1e-47; Mismatches 173; Indels 1; Gaps 1;
QY	6111 tacgagctccagattcaca-tcgtctcttctggtgattgataaaagagaacacgatga 6169
DB	1 tatgaactgccagattcacacgcctctcttggaagattcgaaacagagagaacacgatga 60
QY	6170 tccaagtcccttattctatgatatgacacacaggtgaaacagcgaaattcgcgcaaac 6229
DB	61 tcctgtccatattcttcttccataggaccaggtgaaactgcacatcgcgatgc 120
QY	6230 gcctgaacagaagtgaggagaaagcgtctgcgcaaaatgtctttgacgagactgttc 6289
DB	121 ccccaagacattctgtgattccaggggagacggttagactatgtggaagtccaacatgctt 180
QY	6290 tgaggttaacagctgagggaagcaaaactcacagacagcttcgagggaactcttctgatacc 6349
DB	181 tagttgcaacaataatacagagaatgcaggctcagaagaatcagaggaacacttttgatacc 240
QY	6350 ttgtcgaactgcctatgagagaaagtcttcgcgtcaacggagacatatattccaagtcaacga 6409
DB	241 atgcgcgaacgaatgagaggaagcttccacacttaatggagcgtattttcaagttcaatga 300
QY	6410 gttatttgcagaccagcagtcagctcacaacccatcgatgttctctagagattgatatg 6469
DB	301 ggtatttgcgtgacctattgctcaagtcacaaatccaattgattgctccacgaagtgtgattg 360
QY	6470 ggatctcccaagaaggactgtttacttcggaacatcagtaacatcaatattcagaaggtct 6529
DB	361 ggacctcccaagacgaaactgtttactttggaacctcagttctctacaatattcagaaggttt 420
QY	6530 tccaacggagcagatatacagttctcttcttttggaaaggattcgtatgtcgtggattcga 6589
DB	421 aacgactgaagagatacaacagatgcttttggagaggatttgttgcgtgaggggctttga 480
QY	6590 acagaagacaagagcagcgcgtccatttaattggcaagg 6626
DB	481 taggacagtcgaggcaccacgaaggccctttatgcaagg 517

CC chromosomal DNA in the cell, reduction of expression of the protein in
CC plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDea gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DMEYER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence encodes a DMF-like protein.

XX

SQ Sequence 595 BP; 192 A; 137 C; 134 G; 132 T; 0 other;

Query Match 2.8%; Score 191.8; DB 24; Length 595;
Best Local Similarity 60.4%; Pred. No. 2-2e-38;
Matches 316; Conservative 0; Mismatches 207; Indels 0; Gaps

Qy 6121 cagattcacatcgctccttgatggtgatgataaaagacacagatgatccaagtctt 6180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 cacatccactttacaacagctgaggacttgaccaacgggaacatgatctaccocat 80

Qy 6181 atctcttagtatgacacaccagggtgaacacagcaattcggcacacacgcctgaacaga 6240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 acttatggccatatgacacacagatggaataaaggaaaataactaagacaccacaaacct 140

Qy 6241 agtgtgaggaaaaagctctgcgcaaaatgctgttgacgagacttgtctgagtgtaaca 6300
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 gctgtagccctcaaattgggaggcgatttagcaataatgaaatgtgccacaattgtactg 200

Qy 6301 gtctgagggaacaaactcacagacagttcgcaggaaactcttcgtatccttgcggactg 6360
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 cagsgaaagaaaaccaaactcagatgatactgagaggcacaattctggtctctgcgaacag 260

Qy 6361 ccattgagagaagatttcccgtccaacgggacatatctccaagtccaacgagttattgacag 6420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 ctatgaggggtagttccccacttaatggccacttacttccaagtccaatgagttattgctg 320

Qy 6421 accacgagtcocagtctcaaaccccatcgatgtctcagagattggatggatctcccaa 6480
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 accacagatctagccacaacccaatccatgtgaaaggagagatgctatggaactgcaaa 380

Qy 6481 gaaggactgtttaacttcggaaacatcagtaacataatcattcagaggtcttccaacggagc 6540
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 ggccatggctcttttcgggacttcagtacccaccatttccaaggtctaagacacagaag 440

Qy 6541 agatcacagttcgtctttgaaaggattcgtatgttccgtgatttcgacacagaagacaa 6600
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 aaatcacacaatgctcttgagggggatttctgtgtgcgagatttcgacatggagacta 500

Qy 6601 gagcacgcgtccaatatggcaaggttgcatttctcgcgag 6643
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 qagcaccacagcctctgtgccccatttgcacgttatagcaag 543

Search completed: September 27, 2002, 10:16:06
Job time: 15850 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2002, 05:15:36 ; Search time 4758.11 seconds
(without alignments)
19496.097 Million cell updates/sec

Title: US-09-840-743-5
Perfect score: 6873
Sequence: 1 gttctcggcattgactcgc.....aaaaaaaaaaaaaacctcgag 6873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query		Description	
Result No.	Score	Match	Length	DB	ID		
1	538	7.8	538	9	AV554948	AV554948	
2	533	7.8	533	9	AV554762	AV554762	
3	528.4	7.7	623	9	AV545918	AV545918	
4	525.4	7.6	527	9	AV556668	AV556668	
5	493.6	7.2	500	12	B60854	B60854	
6	478.6	7.0	486	12	B28303	B28303	
7	477	6.9	575	9	AV545791	AV545791	
8	410	6.0	759	10	BG123849	BG123849	
9	403.4	5.9	822	12	BH464280	BH464280	
10	351.6	5.1	657	9	AW267923	AW267923	
11	351.2	5.1	654	10	BG594366	BG594366	
12	326.4	4.7	923	9	BF275260	BF275260	
13	321.6	4.7	923	9	BE035970	BE035970	
14	319.6	4.7	602	9	AW737153	AW737153	
15	318.6	4.6	684	10	BG450287	BG450287	
16	312.2	4.5	537	9	AW092297	AW092297	
17	300	4.4	535	10	BF053858	BF053858	

c	18	299.6	4.4	807	12	BH464289	BH464289
	19	286.4	4.2	781	10	BI308416	BI308416
	20	282.4	4.1	453	10	BI137955	BI137955
	21	281.2	4.1	495	9	AW217654	AW217654
	22	278.2	4.0	515	10	BE433536	BE433536
	23	274.8	4.0	587	10	BF632833	BF632833
	24	265.8	3.9	791	10	BF275262	BF275262
	25	263	3.8	497	10	BE434218	BE434218
	26	262.4	3.8	264	9	BE529258	BE529258
	27	258.4	3.8	644	9	AW686742	AW686742
	28	256.4	3.7	414	10	BE524316	BE524316
	29	246.8	3.6	596	10	BE446194	BE446194
	30	240	3.5	240	12	CNS00S8C	CNS00S8C
	31	232.6	3.4	492	9	AL388226	AL388226
	32	228.8	3.3	665	9	AV833043	AV833043
	33	228.2	3.3	517	10	BE511860	BE511860
	34	228	3.3	228	12	CNS00S8M	CNS00S8M
	35	225	3.3	478	10	BE461624	BE461624
c	37	219.6	3.2	583	9	AV548046	AV548046
	38	214.2	3.1	469	10	BE400675	BE400675
	39	213.6	3.1	487	10	BE248618	BE248618
	40	212.4	3.1	589	9	AV567183	AV567183
	41	211.8	3.1	404	10	BI136405	BI136405
	42	208.4	3.0	865	10	BG416023	BG416023
	43	207.2	3.0	446	10	BE806081	BE806081
	44	197.8	2.9	671	10	BF291322	BF291322
	45	197.4	2.9	524	10	BE515735	BE515735

ALIGNMENTS

RESULT 1

AV554948 AV554948 538 bp mRNA linear EST 06-SEP-2000
LOCUS Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION cDNA clone RZL05C12R 5', mRNA sequence.

ACCESSION AV554948

VERSION AV554948.1 GI:8726363

KEYWORDS EST

SOURCE thale cress

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 538)

AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.

TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and

size-selected cDNA libraries

JOURNAL DNA Res. 7, 175-180 (2000)

MEDLINE 20363093

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

1..538

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="RZL05C12R"

/clone_lib="Arabidopsis thaliana roots Columbia"

/tissue_type="roots"

/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 208 a 102 c 127 g 101 t

ORIGIN

Query Match 7.8%; Score 538; DB 9; Length 538;

```
Best Local Similarity 100.0%; Pred. No. 5.5e-78;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3088 ctgaggttagaatacagatgacccaactgacgagcaagaaaggttaaaaaactg 3147
Db 1 CTGAGGTTATAGAAATCGAGATGATCCAACTGATGGGCAAGAAAGGTAAATAACTG 60
QY 3148 ccagcatcagtaaggtgcactctaaagaaactcgtctccagttataaaagacagcagaaa 3207
Db 61 CCAGCATCAGTAAGGTGCATCTAAAGGAACTCGTCTCCAGTTAAAGACACAGCAGAAA 120
QY 3208 agagagaatattgttccccaaaacgcctgcataaaaggggtcgagcaggttagaaaaaat 3267
Db 121 AGGAGAAATGTATTGTCCCAAAACGCTGCATAAAAGGGTCGAGCAGGTAGAAAAAAT 180
QY 3268 cagtacctcgcctgctcatgcctcagatgacagcttttgcaacctactcctccaaaga 3327
Db 181 CAGTACCTCCGCTGCTCATGCCCTCAGATCCAGATCCAGCTTTGGCAACCTACTCCTCCAAAGA 240
QY 3328 cactttatcaagaagcaagcctaaagaaagggagagaaagttccatacaagattcaggaa 3387
Db 241 CACCTTTATCAAGAAGCAAGCCTTAAGGAAGGAGAGAAAGTCATACAGATTCAGGAA 300
QY 3388 aagcaagaggttccatcaggagaactctgtgtcaggattctattgcggaaataatttaca 3447
Db 301 AAGCAAGAGGTCCATCAGGAGAACTTCTGTCTCAGGATTCTATTGCGGAAATAATTACA 360
QY 3448 ggatgcaaatctgtatctaggagacaaaagaaagagacaaagcaaaatgcaatgtct 3507
Db 361 GGATGCAAAATCTGTATCTAGGACAAAGAAAGAGAAAGCAAAATGCAATGTCT 420
QY 3508 tgtacaagagagatggtgcactttccctatgagagcaagcaagcaagcaagaccca 3567
Db 421 TGTACAAGAGGATGGTGCACTTGTCTATGAGAGCAAGAGCAAGCAAGCAAGACCCA 480
QY 3568 aagttgacattgacgtgaacaactcgcataatggaacttactgaggga 3620
Db 481 AAGTTGACATTGACGATGAAACAACACTCGCATATGGAACCTTACTGATGGGGA 538

RESULT 2
AV554762 AV554762 533 bp mRNA linear EST 06-SEP-2000
LOCUS AV554762 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION cDNA clone RZL01d12R 5', mRNA sequence.
ACCESSION AV554762
VERSION AV554762.1 GI:8726176
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 533)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 533
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL01d12R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"

FEATURES
source
Location/Qualifiers
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 533
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL01d12R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"

Query Match 7.8%; Score 533; DB 9; Length 533;
Best Local Similarity 100.0%; Pred. No. 3.6e-77;
Matches 533; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 206 a 102 c 124 g 101 t
ORIGIN
XhoI"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

QY 3088 ctgaggttagaatacagatgacccaactgacgagcaagaaaggttaaaaaactg 3147
Db 1 CTGAGGTTATAGAAATCGAGATGATCCAACTGATGGGCAAGAAAGGTAAATAACTG 60
QY 3148 ccagcatcagtaaggtgcactctaaagaaactcgtctccagttataaaagacagcagaaa 3207
Db 61 CCAGCATCAGTAAGGTGCATCTAAAGGAACTCGTCTCCAGTTAAAGACACAGCAGAAA 120
QY 3208 agagagaatattgttccccaaaacgcctgcataaaaggggtcgagcaggttagaaaaaat 3267
Db 121 AGGAGAAATGTATTGTCCCAAAACGCTGCATAAAAGGGTCGAGCAGGTAGAAAAAAT 180
QY 3268 cagtacctcgcctgctcatgcctcagatgacagcttttgcaacctactcctccaaaga 3327
Db 181 CAGTACCTCCGCTGCTCATGCCCTCAGATCCAGATCCAGCTTTGGCAACCTACTCCTCCAAAGA 240
QY 3328 cactttatcaagaagcaagcctaaagaaagggagagaaagttccatacaagattcaggaa 3387
Db 241 CACCTTTATCAAGAAGCAAGCCTTAAGGAAGGAGAGAAAGTCATACAGATTCAGGAA 300
QY 3388 aagcaagaggttccatcaggagaactctgtgtcaggattctattgcggaaataatttaca 3447
Db 301 AAGCAAGAGGTCCATCAGGAGAACTTCTGTCTCAGGATTCTATTGCGGAAATAATTACA 360
QY 3448 ggatgcaaatctgtatctaggagacaaaagaaagagacaaagcaaaatgcaatgtct 3507
Db 361 GGATGCAAAATCTGTATCTAGGACAAAGAAAGAGAAAGCAAAATGCAATGTCT 420
QY 3508 tgtacaagagagatggtgcactttccctatgagagcaagcaagcaagcaagaccca 3567
Db 421 TGTACAAGAGGATGGTGCACTTGTCTATGAGAGCAAGAGCAAGCAAGCAAGACCCA 480
QY 3568 aagttgacattgacgtgaacaactcgcataatggaacttactgaggga 3620
Db 481 AAGTTGACATTGACGATGAAACAACACTCGCATATGGAACCTTACTGATGGGGA 533

RESULT 3
AV545918/c AV545918 623 bp mRNA linear EST 06-SEP-2000
LOCUS AV545918 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION cDNA clone RZL05C12F 3', mRNA sequence.
ACCESSION AV545918
VERSION AV545918.1 GI:8717332
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 623)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
```

source 1. .523
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="R2L05c12F"
 /clone_lib="Arabidopsis thaliana roots Columbia"
 /tissue_type="roots"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 177 a 127 c 135 g 184 t
 ORIGIN

Query Match 7.7%; Score 528.4; DB 9; Length 623;
 Best Local Similarity 99.6%; Pred. No. 1.9e-76;
 Matches 540; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6312 gcaaaactcacagacttcgagaaactcttcgatacttgcgagactgccatgagagga 6371
 |||
 Db 623 GCAAACTCACAGACAGTTCGAGAACTCTTCGATACCTTGTGCGAGCTGCCATGAGAGGA 564
 |||

QY 6372 agtttcgcgtcaacgggacatattcccaagtcacagagattattcgaccacagagtc 6431
 |||
 Db 563 AGTTTCCGCTCAACGGGACATATTTCCAACTCAACGAGTTATTTTCGACACACAGTCC 504
 |||

QY 6432 agtctcaaaacctcgatgttcctgagagattggatggatctcccaagaggactgtt 6491
 |||
 Db 503 AGTCTCAAAACCCATCGATGTTCTAGAGATTGGATATGGGATCTCCCAAGAGGACTGTT 444
 |||

QY 6492 tacttcggaacatcagtaacataatattcagaggtctttcaacggagcagatacagttc 6551
 |||
 Db 443 TACTTCGGAACATCAGTAACATCAATATTTCAGAGGTCTTTTCAACGGAGCAGATACAGTTC 384
 |||

QY 6552 tgcctttggaaagattcgtatgtccgtgattcgacagagaagacagaccgcgt 6611
 |||
 Db 383 TGCCTTTGGAAGGATTCGTATGTGTCCTGGATTCGACACAGAACACAGACCCCGGT 324
 |||

QY 6612 ccattaatggcaaggttgcatcttcctgcgagcaaatggaagacaacaaacccataaga 6671
 |||
 Db 323 CCATTAAATGGCAAGTTGCAATTTCTCGGAGCAAAATTTGAAGAACAAACCACTTAAAGA 264
 |||

QY 6672 tgactggaagaagcaaacgcattgtctctcgtctcctctatttaagccaggaaag 6731
 |||
 Db 263 TGACTGGAAGAAAGCAACGCAATGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 204
 |||

QY 6732 tccatttagacataataacaggaatccaaataggctatttctcttctcttcttcttctt 6791
 |||
 Db 203 TCCATTAGACATATATACAGAAATCCAAATAGGCTATTTTCTCTCTCTCTCTCTCTCT 144
 |||

QY 6792 attcatagagcagaagcgacac-aaaaaagtttttgggtatttatttctctcttaaca 6850
 |||
 Db 143 ATTCTATAGACAGACGACACAAAAAAGTTTTTTGGGTTATTATTCTCTCTCTCTCTCT 84
 |||

QY 6851 aa 6852
 ||
 Db 83 AA 82

RESULT 4
 AV556668
 LOCUS B60854/c
 DEFINITION T1907TF TAMU Arabidopsis thaliana genomic clone T1907, DNA sequence.
 ACCESSION B60854
 VERSION B60854.1
 KEYWORDS GI:2627785
 SOURCE GSS.
 ORGANISM Arabidopsis thaliana
 thale cress.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 527)
 AUTHORS Asamizu,Y., Nakamura,Y., Sato,S. and Tabata,S.

SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 500)
Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
,J.C.
TITLE
A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3
JOURNAL
COMMENT
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 500.
Location/Qualifiers
1. .500
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T1907"
/clone_lib="TAMU"
/sex="hermaphrodite"
/notes="Vector: BelOBAcII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
BASE COUNT 159 a 138 c 70 g 133 t
ORIGIN

Query Match 7.2%; Score 493.6; DB 12; Length 500;
Best Local Similarity 99.2%; Pred. No. 9.6e-71;
Matches 496; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 396 caaaaaatcattatgcatattatggtgcagattcttaattatgctctcttaac 455
|||||
Db 500 CAAAAATCATTATTGTCATTATGTCGCAGATTCTTAGTTAATGTCGCTTCTTAAC 441
|||||
Qy 456 caagtcagattaaaaagaggttcgtccatgttgccttcttgggttggagagagt 515
|||||
Db 440 CAAGTCAGATTAAAGAGGTGTCGTCCTGTCGTTGTTGGTGTGGAGAGGT 381
|||||
Qy 516 ttccgagagattggtgagtgatttatttgggtgagtagtgatgaagtttgaaggggag 575
|||||
Db 380 TTTCCGAGAGCTTAGGTGAGTGTATTTTGGGTGAGGTAGTGAAGGTTTGAAGGGGGAG 321
|||||
Qy 576 tgattcatcaagtgattatgaattcagaggttcgatccgggggatagatatatttcagtt 635
|||||
Db 320 TGATTTCATCAAGTGTGTATGAATTCAGGGCTGATCCCGGGGATAGATATTTCGAGTT 261
|||||
Qy 636 cctttggagaatcaaacacaaagagttcattgggttccttgattccattacacccaaa 695
|||||
Db 260 CCTTTGGAGAAATCAAACTCAACAAGAGTTCATGGGTCTTGGATTCCATTACACCCAAA 201
|||||
Qy 696 aaacctagatcaagtcgtatgtagatgagagtgatgaataaccagagatcctaaatgggttt 755
|||||
Db 200 AAACCTAGATCAAGTCTGTAGTGTAGATGAGAGAGTGTATAAACACGAGGATCTAAAATGGGTTT 141
|||||
Qy 756 ccaggttggtgaattttagacagggattctcaaacactggttgatcataatgggtt 815
|||||
Db 140 CCAGGTGGTGAATTGTAGACAGGGGATCTGCAACHACTGGTGTGGATCAATAATGGGTTT 81
|||||
Qy 816 ttgatcatggtgtctcatcagggcggtttaccacacttaagtatgatgataatagcttagcg 875
|||||
Db 80 TTTGATCATGCTCATCAGGGCGGTTTACCAACTTAAGTATGATGATCAATAGCTTAGCG 21
|||||
Qy 876 ggaatcacatgcacaaagcttg 895

|||||
Db 20 GGATCACATGCACAGCTTG 1
|||||

RESULT 6
B28303
LOCUS
DEFINITION
B28303 TAMU Arabidopsis thaliana genomic clone T9J15, DNA
sequence.
486 bp DNA linear GSS 13-OCT-1997
ACCESSION B28303
VERSION B28303.1 GI:2514269
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 486)
Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and
Venter,J.C.
TITLE
Use of a BAC End Sequence Database To Identify Minimal Overlaps for
Arabidopsis Genomic Sequencing
JOURNAL
COMMENT
Other_GSSs: T9J15TFB
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 486.
Location/Qualifiers
1. .486
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T9J15"
/clone_lib="TAMU"
/sex="hermaphrodite"
/notes="Vector: BelOBAcII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
BASE COUNT 145 a 96 c 106 g 138 t 1 others
ORIGIN

Query Match 7.0%; Score 478.6; DB 12; Length 486;
Best Local Similarity 99.0%; Pred. No. 2.7e-68;
Matches 481; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1120 caagcttgctgagtggtgggatagcttattcaaggttcgctcaatgagtgatgataatc 1179
|||||
Db 1 CCAGCTGCTGAGTGGGGGATAGCTTATTCAGGTTCTGTCATGTCGAGTGAATCAATC 60
|||||
Qy 1180 tatttcagttttttttttcccttttccgtttcttgcagtcacttagagtagaacaatga 1239
|||||
Db 61 TATTTTCAGTTTTTTTTTCCCTTTCTTCGTTCTTCGAGTACTTAGTAGTAGATCATGA 120
|||||
Qy 1240 attagaatacttaagaagaatcattggtttgaacagatggacctccagcgtgtaacaagc 1299
|||||
Db 121 ATTAGAATATCTTAAGAAAGTCATGTTTTTGAACAGATGGACCTCCACGCGTGAACAAGC 180
|||||
Qy 1300 cttttacaatttgaattccaccattagaagaagcagttgggtcagttctgtgaaagt 1359
|||||
Db 181 CTCTTTACAAATTTGAATTCACCAATTTAGAGAGAGAGCAGTTGGGTCACTCTGTGAAAGTT 240
|||||
Qy 1360 cggttcaatgtacogtcaacgcccgctgtttcagaacagggtgaaaaactggattcc 1419
|||||
Db 241 CGTTTCAATATGTACCGTCAACGCCAGCTCTGTTTCAGAACAGGTGAAAAGACTGGATTCC 300
|||||
Qy 1420 ttgaacagatagttacaactacttgacatgaaatccccagagcgaaatctgacaaaaagta 1479


```

Db 301 TTTAAACAGATAGTTACAACTACTGGACATGAATCCAGAGCGGAGATCTGACAAAGTA 360
QY 1480 tgcagagcattatggactcgtcgtctgtaatacgacggaagctactgaacaaatgatg 1539
Db 361 TGCAGAGCATTATGGACTCGTCTGCTGTTAATGCCAGCGGAAGCTACTGAACAAATGATG 420
QY 1540 gcagcagacaagatgtctcgtgagtgacaccttaacaaactcctcagcagaacccctcca 1599
Db 421 GCAGCAGACAAGATGTTCTGGAGTTGCGACCTTAACAAAACTCCTCTGCAGAACCCCTCCA 480
QY 1600 aaagga 1605
Db 481 NAAGGA 486

RESULT 7
LOCUS AV545791/c 575 bp mRNA linear EST 06-SEP-2000
DEFINITION AV545791 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
cDNA clone R2L0Ld12F 3', mRNA sequence.
ACCESSION AV545791
VERSION AV545791.1 GI:8717205
KEYWORDS EST
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 575)
AUTHORS Asanizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
source
location/Qualifiers
1..575
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="R2L0Ld12F"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 167 a 116 c 119 g 173 t
ORIGIN

Query Match 6.9%; Score 477; DB 9; Length 575;
Best Local Similarity 99.8%; Pred. No. 4.7e-68;
Matches 488; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6365 gagaggaagtcttcgctcaacgggacatatattccaagtcacagagttatttgcagacca 6424
Db 575 GAGAGGAAGTTTTCGCTCAACGGGACATATTTCCAAGTCAACGAGTATTATTCGACACCA 516
QY 6425 cgagtcagctcacaacccatcgatgtctcctagatgtgatatggatctcccaagaag 6484
Db 515 CGAGTCCAGCTCCTCAACCCCATCGATGTTCTTAGAGATTGGATGGATCTCCCAAGAAG 456
QY 6485 gactgttacttcggaacatcagtaacaataftcagaggtctttcaacggagcagat 6544
Db 455 GACTGTTTACTTCGGAACATCAGTAACATCAATATTCAGAGGCTCTTCAACGGAGCAGAT 396
QY 6545 acagttctcgttttggaaaggattcgatgtgtcctcgttgattcgaacagaagacagagc 6604

```

```

Db 395 ACAGTTCCTCTTTTGGAAAGGATTGCTATGTGTCCGTGGATTGGAACAGAAAGAGC 336
QY 6605 acggttcattaaatgcaaggttgcatcttctcgcagcaaatgaagaacaaac 6664
Db 335 ACCGGCTCCATTAAATGGCAAGGTTCGATTTTCTCGGACAAATTCGAACAAACAAAC 276
QY 6665 ctaagatgactgggaagaagaacgcattgtctctcgtctcctcctcctatttaagcca 6724
Db 275 CTAAGATGACTGGGAAGAAGCAACGCAATGCTTCTCTGCTCTCTCTATTATTAAGCCA 216
QY 6725 ggaagagtcctcatttagacataataacgaagaatccaaatagactatttcttcttcttc 6784
Db 215 GGAAGATCCCATTTAGACATAATAACAGGAATCCAAATAGGCTATTTCTCTTTCTTC 156
QY 6785 ttatttcattcatagagcagagcgacac-aaaaagttttttgggtatttatttttctc 6843
Db 155 TTATTTTCATTATAGAGCAGAGCGACACAAAAAAGTTTGGGTATTATTTTCTC 96
QY 6844 tctaacaaa 6852
Db 95 TCTAACAAA 87

RESULT 8
LOCUS BG123849
DEFINITION BG123849 759 bp mRNA linear EST 31-JAN-2001
EST469495 tomato shoot/meristem Lycopersicon esculentum cDNA clone
CTOP3G24 5' sequence, mRNA sequence.
ACCESSION BG123849
VERSION BG123849.1 GI:12624037
KEYWORDS EST
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 759)
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
Hansen,C., Ronning,C. and Tanksley,S.
TITLE Generation of ESTs from tomato shoot/meristem tissue
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
source
location/Qualifiers
1..759
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="ctOP3G24"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Small expanding leaves from the growing tip were
taken from greenhouse plants (4-6wks old TA496). Tissue
was immediately frozen in liquid nitrogen."
BASE COUNT 225 a 158 c 191 g 185 t
ORIGIN

Query Match 6.0%; Score 410; DB 10; Length 759;
Best Local Similarity 73.4%; Pred. No. 3.5e-57;
Matches 524; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 5062 tcagaaaaatgtggagggaatgaaggagagacaggaacaaacaaacaatatagatt 5121
Db 4 TGAGAAAGGAAGTCCCAATCAAGAGTGGGAAAAAAGAAAGCAAGGATGCAATGGACT 63

```



```
tubers."
BASE COUNT      192 a   148 c   157 g   157 t
ORIGIN

Query Match      5.1%; Score 351.2; DB 10; Length 654;
Best Local Similarity 74.8%; Pred. No. 1.4e-47;
Matches 440; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 5129 ctatgaacataagagctgctagtatcatcagcgagattctgagcgtatcaaggaagagg 5188
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 CTGGACGCACTGAGGTGTGCAGATGTCATAGATAGCCCATACCATAGAGAACGGGG 61
Qy 5189 gatgaataacattgtggtgacgaattaaagatttctgaacagcgatagttaaagatca 5248
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 AATGAACAATATGTAGCAGAGAGAAATCAAGGATTTCTTAATCGGATTTTGTAGAGAACA 121
Qy 5249 tgggtggtatcgacctgaattggttgagagaatctctctctgataaagccaaggactatct 5308
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 TGGAGATATTGATCTTGATGGCTGAGAGACGTTCCACGACAGACAAAGCAAAAGAGTATCT 181
Qy 5309 cttagcataagagctggtgttgaaagtgttgaaatgctgagactcttaacactcca 5368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 ATTAGCAATCTGGGGCTTGGCCCTGGAAGAGCGTAGAGTGTGTGGCACTTTTAACACTTCA 241
Qy 5369 caattcttctccctgttgacagaaatgttggaaaggtagcagttaggatgggtgggt 5428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 CAACCTTGCTTTTCTGTCGACACAAATGTTGGCGGTATAGCTGTGTAGACTGGGATGGGT 301
Qy 5429 gcctctacaacccctacatgaatcaactcactcagttacacactcctgagctataccagtgct 5488
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 GCCCTGTCAGCCACTGTCAGAGAGTCGCTACAATGTCATCTTCTGGAACGTGTACCCCTGTGCT 361
Qy 5489 cgaatccatcaaaaattcttggccaagacttggcaactcgatcaacgaacactgta 5548
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 GGAGTCAATTCAMAAATACCTCTGCCCACGCTCTGCAAGCTTGTATCAAGAAGACATTATA 421
Qy 5549 tgaattacactaccaactgattacgttttggaaaggtattttgacaaagagtagaccaaa 5608
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 TGAGTTACATTACCATATGATGATCACTTTTGGAAAGGCTCTCTGTTCAAAAGCAAAACCCAA 481
Qy 5609 ttgtaatcatgtccaatgagagagagtgagacaacttggccagtgcttatgtagtgc 5668
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 CTGCAACGCTGTCCAAATGAGAGAGAGATGTCAGACACTTCCGCAAGTGTCTTTGCAAGTGC 541
Qy 5669 aagactctgttaccgcaacagagagagagagcttaacaagtgcac 5716
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 AAGCTTGCCCTGCCACGACCAAGAGAGAGATATAGTGTAGTGCCAC 589

RESULT 12
LOCUS      BF275260
DEFINITION GA_EB0023J04f Gossypium arboreum 7-10 dpa fiber library Gossypium
           arboreum cDNA clone GA_EB0023J04f, mRNA sequence.
ACCESSION  BF275260
VERSION     BF275260.1
KEYWORDS   EST.
SOURCE     Gossypium arboreum.
ORGANISM   Gossypium arboreum.
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE  1 (bases 1 to 768)
AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
           D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE     An integrated analysis of the genetics, development, and evolution
           of the cotton fiber
JOURNAL   Unpublished (2000)
COMMENT   Contact: Wing RA
           Clemson University Genomics Institute
           Clemson University
           100 Jordan Hall, Clemson, SC 29634, USA
```

QY 6526 gtctttt 6531
 | | | | |
 Db 751 GACTTT 756

RESULT 13
 BE035970
 LOCUS MO22F11 MO Mesembryanthemum crystallinum cDNA 5' similar to protein
 DEFINITION , mRNA sequence.
 ACCESSION BE035970
 VERSION BE035970.1 GI:8330979
 KEYWORDS EST.
 SOURCE common ice plant.
 ORGANISM Mesembryanthemum crystallinum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Caryophyllales; Caryophyllales; Alzooceae; Mesembryanthemum.
 REFERENCE 1 (bases 1 to 923)
 AUTHORS Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
 ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
 Scara,G., Wheeler,M. and Zepeda,G.R.
 TITLE Functional Genomics of Plant Stress Tolerance
 JOURNAL Contact: Michalowski,C.B.
 COMMENT University of Arizona
 Bio Sciences West room 513, Tucson, AZ 85721, USA
 Tel: 520-621-7982
 Fax: 520-621-1697
 Email: cbm@u.arizona.edu
 An open reading frame exists.
 Location/Qualifiers
 1. 923
 /organism="Mesembryanthemum crystallinum"
 /db_xref="taxon:3544"
 /clone_lib="MO"
 /tissue_type="apical meristem and leaf primordia"
 /dev_stage="5 weeks"
 /note="no stress"

BASE COUNT 290 a 216 c 202 g 214 t 1 others
 ORIGIN

Query Match 4.7%; Score 321.6; DB 9; Length 923;
 Best Local Similarity 65.6%; Pred. No. 8.4e-43;
 Matches 520; Conservative 0; Mismatches 260; Indels 13; Gaps 3;

QY 5816 tagaataacttgacccaataattgaagagccgctgcgcggggaagagtgcaactga 5875
 | | | | |
 Db 59 TGGAAATTACTTGTGAGCCAAATTTGTGAGTTCGGCATCACCAGAACCCAGAG--CAGGA 115
 | | | | |
 QY 5876 aataacggagagtattgaagagcttaactacaatgagagccctgacagatcccaac 5935
 | | | | |
 Db 116 ACAACATATTCCGACATTTGAAGAT-----TATGAAGATCTGTGATGAATTTCCCAAC 166
 | | | | |
 QY 5936 aataaactcaacattgaacagtttggatgactctacgggaacacatggaagaacacat 5995
 | | | | |
 Db 167 AATCAAGCTCAATATGAAGAATTTCACTCAGACACTAGAGAAATTTATGCAAGGAATAG 226
 | | | | |
 QY 5996 ggagtcacaagaagtgacatgtccaaagcttggctgttgcattccaaacactacttc 6055
 | | | | |
 Db 227 AGAACTTCAAGAGACTGCACCTATCAAGGGCCCTTGGTGGCTTTAACACCTGGAAGCTGCTTC 286
 | | | | |
 QY 6056 tattccaactcccaactaaagaacattagcgtctcagcagacgacccaagtacga 6115
 | | | | |
 Db 287 CATTCCTACACCAAGCTCAAGAAATATACGTCCGCTAGAACAGAGCATCATGTCTACGA 346
 | | | | |
 QY 6116 gtccacagattcacatcgctcctcttgatggatgataaaagacacagatgacccaag 6175
 | | | | |
 Db 347 ACTTCCGACTCACACCCTCTTTTAAAGGATTGGCAACAGAGACCTGATGATCCGTG 406
 | | | | |
 QY 6176 tccctatctcttagtatatgacacacaggtgaaacagcgaattcggcacaccgcctga 6235
 | | | | |

Db 407 TTCATACCTACTTGCAATCTGGACCCCGAGACCCGAAATTCATCCAACTCCTGA 466
 | | | | |
 QY 6236 acagaagtgtgagggaaagcgtctgcgcaaaatgtgctttgacgagactttcttgagt 6295
 | | | | |
 Db 467 AAGAAGATGCAGCTCTGAAGACCCAAACTTACTTTGCAATGAGCAACATGCGCCTACTG 526
 | | | | |
 QY 6296 taacagctgagggagaaacacacagacagcttcgaggaactctcttgatacccttgcg 6355
 | | | | |
 Db 527 TAGCAGTCAGCGGGAAGTAACCTCCAAATTTGTGGNGAACTCTTCTGATACCATGAG 586
 | | | | |
 QY 6356 gactgccaatgagagaaagtcttcgctcaacgggacacattttcaagtcacagcaggtatt 6415
 | | | | |
 Db 587 AACAGCAATGAGAGGAGCTTTCCACATCAATGGAACATACTTTCAAGTAAACGAGGTGT 646
 | | | | |
 QY 6416 tgcagaccacagctccagctctcaaacccatcgatgttccctagagattgatatggatct 6475
 | | | | |
 Db 647 TGCAGACCATGACTCCACCCCTTACCCAGTGGAGTGCCACAGATCTTTGTGTGGAACCT 706
 | | | | |
 QY 6476 cccaagaaggactgttacttcggaacatcgatcaatcaatattccagagtgcttcttcaac 6535
 | | | | |
 Db 707 GCCAAGACGAACCTGTATA-TTTTGTACTTCTATATACCCACCATATTACAGAGCTGAACAC 765
 | | | | |
 QY 6536 ggagcagatacagttctcttggaaaggattcgatgtcgcgtgagtcgcgaacagaa 6595
 | | | | |
 Db 766 TCAAGATATTCAACATTCGTTTTTGGAGAGGTATGTTTTCGTTAGAGGGTTTGACAAAA 825
 | | | | |
 QY 6596 gacaagagcacgcg 6608
 | | | | |
 Db 826 AGAGAGACCCCG 838

RESULT 14
 AW737153
 LOCUS EST338580 tomato flower buds, anthesis, Cornell University
 DEFINITION Lycopersicon esculentum cDNA clone cTOD2A6 5', mRNA sequence.
 ACCESSION AW737153
 VERSION AW737153.1 GI:7646098
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 REFERENCE 1 (bases 1 to 602)
 AUTHORS van.der Hoeven,R.S., Bezzeredes,J.L., Matern,A.L., Holt,I.E., Liang
 ,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman
 ,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
 Generation of ESTs from tomato flower tissue, anthesis
 Unpublished (1999)
 TITLE Contact: CUGI
 JOURNAL Clemson University Genomics Institute
 COMMENT 100 Jordan Hall, Clemson, SC 29634, USA
 Email: http://www.genome.clemson.edu/orders/index.html
 5 prime sequence.

FEATURES
 Location/Qualifiers
 1. 602
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cTOD2A6"
 /clone_lib="tomato flower buds, anthesis, Cornell
 University"
 /tissue_type="flower"
 /dev_stage="anthesis"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Tanksley; Flower buds and flowers were
 taken from greenhouse plants (4-8 wks old, TA496). They
 were immediately frozen in liquid nitrogen and then
 size-separated while remaining frozen."

BASE COUNT 158 a 136 c 151 g 157 t

Search completed: September 27, 2002, 07:45:12
Job time: 8976 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 10:26:26 ; Search time 70.37 Seconds
(without alignments)
2360.927 Million cell updates/sec

Title: US-09-840-743-2
Perfect score: 9089
Sequence: 1 MQSIMDSSAVNATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3949	43.4	1017	T48452	hypothetical prote
2	2863	31.5	555	T48453	hypothetical prote
3	1559.5	17.2	1207	D84781	hypothetical prote
4	1260	13.9	234	T48454	hypothetical prote
5	709	7.8	917	T05430	hypothetical prote
6	278	3.1	5327	T13564	microtubule-associ
7	246.5	2.7	2897	B48666	cell proliferation
8	246.5	2.7	3256	A48666	cell proliferation
9	240.5	2.6	2938	T30249	cell proliferation
10	233.5	2.6	2218	B84683	hypothetical prote
11	232	2.6	259	D75275	endonuclease iii -
12	219.5	2.4	1871	D96796	probable heat choc
13	217.5	2.4	1490	T20513	hypothetical prote
14	217	2.4	2022	T48818	glucan 1,4-alpha-g
15	211	2.3	1188	T05324	hypothetical prote
16	210.5	2.3	1805	A34736	nestin - rat
17	208.5	2.3	3924	S37431	ankyrin 2, neurona
18	207	2.3	3331	T25410	hypothetical prote
19	206	2.3	3187	JC5837	364K Golgi complex
20	205.5	2.3	2562	T14266	Xin protein - chic
21	205	2.3	1983	T00385	KTA0624 protein -
22	204.5	2.2	2717	A34203	DNA-binding protei
23	197.5	2.2	1145	T18235	transcription acti
24	197	2.2	1366	B86292	F/H2.12 protein -
25	196.5	2.2	3225	I52300	giantin - human
26	196	2.2	3488	T34418	hypothetical prote
27	192.5	2.1	2954	T14156	kinesin-related pr
28	192.5	2.1	3259	A36539	giantin - human
29	192	2.1	1888	T14273	zinc finger protei

30	191.5	2.1	2094	2	S33124	tpr protein - huma
31	191	2.1	1742	2	T49451	kinesin-like prote
32	191	2.1	2052	2	C97038	phage-related prot
33	190.5	2.1	2447	2	T16870	hypothetical prote
34	190.5	2.1	4957	2	T03455	ALR protein - huma
35	190.5	2.1	5262	2	T03454	ALR protein - huma
36	190	2.1	1284	1	HWZAI	A-type inclusion p
37	190	2.1	2464	1	QRMSPI	microtubule-associ
38	189.5	2.1	236	2	S75373	probable DNA-(apur
39	189.5	2.1	1634	2	T26517	hypothetical prote
40	189.5	2.1	1804	2	T34518	nestin - golden ha
41	189.5	2.1	3498	2	T22330	hypothetical prote
42	189	2.1	2453	2	S60234	nuclear receptor c
43	188.5	2.1	1435	2	A37793	erythrocyte-bindin
44	188.5	2.1	1926	2	JC4842	DNA-binding nuclea
45	188	2.1	1593	2	T22028	hypothetical prote

ALIGNMENTS

RESULT 1

T48452

hypothetical protein T32M21.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48452

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24487

A:Accession: T48452

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1017 <BEV>

A:Cross-references: EMBL:AL162875

A:Experimental source: cultivar Columbia; BAC clone T32M21

C:Genetics:

A:Map position: 5

A:Introns: 167/1: 874/1

A>Note: T32M21.160

Query Match 43.4%; Score 3949; DB 2; Length 1017;

Best Local Similarity 99.9%; Pred. No. 1.1e-206;

Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSIMDSSAVNATEQNDGSRQDVLEFDLNKTPOQKPSKRKRKFMKPVVVEGPKRKP 60

Db 235 MQSIMDSSAVNATEQNDGSRQDVLEFDLNKTPOQKPSKRKRKFMKPVVVEGPKRKP 294

QY 61 RPAELPKVVEGPKRPRKRAATQEKVKSTGSAKKNLKESATKKPANVDMSKSP 120

Db 295 RPAELPKVVEGPKRPRKRAATQEKVKSTGSAKKNLKESATKKPANVDMSKSP 354

QY 121 EVTLKSCRKALNFDLENPGDARQDSESEIVQNSGANSFSEIRDAIGGTGNSFLDSVSQ 180

Db 355 EVTLKSCRKALNFDLENPGDARQDSESEIVQNSGANSFSEIRDAIGGTGNSFLDSVSQ 414

QY 181 IDKTNGLGAMNQPLEVSGMGNQDPDKLSTGAKLARDQDPDLLTRNQOCQFPVATQTFPME 240

Db 415 IDKTNGLGAMNQPLEVSGMGNQDPDKLSTGAKLARDQDPDLLTRNQOCQFPVATQTFPME 474

QY 241 NQAAWLQMKNLQIGFPFGNQPMYIRNOQPCLANGNQPMYLGTPRALVSGNQQLGG 300

Db 475 NQAAWLQMKNLQIGFPFGNQPMYIRNOQPCLANGNQPMYLGTPRALVSGNQQLGG 534

QY 301 POGNKRPIFLNHTQCLPAGNOLYGSPTDMHQLVMSTGGQGHLLIKNOQPGSLIRGQQPC 360

Db 535 POGNKRPIFLNHTQCLPAGNOLYGSPTDMHQLVMSTGGQGHLLIKNOQPGSLIRGQQPC 594

QY 361 VPLIDQQPATPKGFTHLNQMVATSMSSPGLRPHSOSQVPTTYLHVESVRLNGTGTGQC 420


```
Db 332 BGIFGQGESLNLVKIDPIPKKTTGHARFNLSSMKNLV-----EV 376
Qy 399 PTYLHVSVSRILNGTTGTCQRAPAYDSLOODIHKGNKYILSHSINGCKKALPQ 458
Db 377 PE---HL-----TSYGCKKQ-----QNNKILVDTRVT-----401
Qy 459 NSSLPPTIMAKLEBARSKROYHRAMGQTEKHDLNLAAQIAQSOODVERHNSSTCVELDA 518
Db 402 -----VSKKPKTKSKSO-----414
Qy 519 AKTKIOKVVQENLHGWPVEIETEDPTDGARKKNTASIKASKGNSPVKKAETKE 578
Db 415 ---TK-QKNLPLNLCRPPSTGLSPD-----437
Qy 579 KCIVPKTPAKKGRAGKKSVPVPAHASEIQWLQWPTPLRSRKPCKGKRSIQDSGKA 638
Db 438 -----ELW-----KRENSI-----446
Qy 639 RCPGSELLCQDSIAEIIYRMONLYLQKEREQONAMVLYK-----GDGALVPE 688
Db 447 -----ETISEL-----LRLLDINREHSETALVPYTMNSQIVLFGGAGAIPTVT 490
Qy 689 S-KKKRPKPKVIDDETRINWLLMGDEKEGDEEDKKKEKWEEERRVFRGRADFSI 747
Db 491 PVKRPKPKVLDODETRVWKKLL-ENINSEGVGGDEQKAKWEEERNVFRGRADFSI 549
Qy 748 ARMHLVQDRRFSPKGSVDVSGVFLTONVSDHLSSAFMSLAARFP-PKLSSSREDE 806
Db 550 ARMHLVQDRRFSPKGSVDVSGVFLTONVSDHLSSAFMSLASQPPVPFSSNFD- 608
Qy 807 RNRSVVVEDPEGCLINLETPSQW-----EKVQHPSDMEVSGVDSGSKQLRDCSNS 859
Db 609 -----AGTSSMPQIITYLQSEETMSPPDHNSVT-----640
Qy 860 GIERNFLEKSIQNLBEEVLSSQDSFDPAIFQSGRVCSCSKSDAEFTTRCET--KT 917
Db 641 -----LKNTPQDEEKDYVPSNET-----SRSSSEIAISAHESVDKT 676
Qy 918 VSGTS--QSVQTGSPNLSDSEICLOGNERPHLYEGSGDVQK-QETTNVAQKKPDLKTMNW 974
Db 677 TDSKEYVDSDRKGS-----SVEVDKTEKCRVLNLFPSDSALTC 716
Qy 975 KDSVCFQCPNDTNWQTPSSSYEQCATROPHVLIDIEFGMQGGLGYSWMSISPRVDRV 1034
Db 717 QHSMVSDAPQNT---ERAGSS-----EIDLEGE-YRTSEMKL-----750
Qy 1035 KKNVPRFRFGGSGVPRFTGTIIPSTPHELPGMGLSGSSSAVQEHODDTQHQQDEMN 1094
Db 751 -----LQGVQVLEDSNQVSPN-----MSPGDCSSSETKGFQ-----SMKEPT 787
Qy 1095 KASHLQKTFDLNLSSE-CLTROSTSKQNTDGCCLPRDRTAEVDVDPPLSNSSLNILV 1153
Db 788 KSS-----VDSSEPGCCSQ-----DG-----DVL-----807
Qy 1154 ESNSSNKEQTAVEYKFNATILREMGKTLADGKKPTQOWDSLKDKVNEGROGRNNNM 1213
Db 808 ---SCOKPT---LKREKGVKYLKE-----EKKAEDWDCLREAAQARAGIREKTRSTM 852
Qy 1214 DSIDYEAIRRASISFISPAIKERGNMNLAVRI-----KDFLERIVKDHGGIDLE 1263
Db 853 DTVDMKAIKRAADVKEVATETIKHSMNHKLAERIQYLTNNKIMQGFGLDLRLVNDHGSIDLE 912
Qy 1264 WLRESPPDKAKDYLLSIRGLKSVCEVRLTLHLNLAFFPVDNTNVGRIAVRMGWVPLQPLP 1323
Db 913 WLROVPPDKAK-----923
Qy 1324 ESLQLHLELXPVLESQKFLWPLRCLKDORTLYELHYQLITTFKGVCTKSRPNCNACPM 1383
Db 924 -----YELHYQMITEFGKVFCTKSKPKNCNACPM 950
Qy 1384 RGEGRHFASAVASARLALPAPEERSLTSATIPVPPEFPPVAIPMIELPPLPLEKSLAGA 1443
Db 951 KGEGRHFASAFASARLALPSTEKGMGTPDKNPLP-----LHLPEPFOREOGSEV 999
```

```
Qy 1444 PSNREN-----CEPIIEEPASPCQECTEITESDIEDAYYNEDPDEIPTIKLNIEQFGMT 1497
Db 1000 VQHSEPAKKVTCCEPIIEEPASPEPEAEVSIADIEEAF-EDPEEPTIRLNMDAFTSN 1058
Qy 1498 LREHMERMELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYVELPDSHRLDGM 1557
Db 1059 LKIMEHNKLDGDNMSSALVALTAETASLPMPKLNISQLRTEHRYVELPDEHPLLAQL 1118
Qy 1558 DKREPDDPSYLLAIWTPGTANSAPPEQKCGKASGKMGCFDETCSECNSLREANSQTV 1617
Db 1119 EKREPDDPSYLLAIWTPGTADSIQPSVSTCFQANGMLCDEETCFSCNSIKETRSQIV 1178
Qy 1618 RGTILLI 1623
Db 1179 RGTILV 1184

RESULT 4
T48454
hypothetical protein T32M21.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48454
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48454
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics.
A:Map position: 5
A:Introns: 49/2; 61/3; 81/1; 127/3; 149/3; 188/1; 201/1
A:Note: T32M21.180

Query Match 13.9%; Score 1260; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.9e-61;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1496 MTLREHMERMELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYVELPDSHRLD 1555
Db 1 MTLREHMERMELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYVELPDSHRLD 60

Qy 1556 GMDKREPDDPSYLLAIWTPGTANSAPPEQKCGKASGKMGCFDETCSECNSLREANSQ 1615
Db 61 GMDKREPDDPSYLLAIWTPGTANSAPPEQKCGKASGKMGCFDETCSECNSLREANSQ 120

Qy 1616 TVRGTLILPCRTAMRGSPFLNGTYFQVNELFADHESLSKPIDVPRDWIWLPRRTVYFGT 1675
Db 121 TVRGTLILPCRTAMRGSPFLNGTYFQVNELFADHESLSKPIDVPRDWIWLPRRTVYFGT 180

Qy 1676 SVTSIFRGLSTEQIOFCFWKGFVVCVRGFEQKTRAPRPLMARLHPPASKLKNKT 1729
Db 181 SVTSIFRGLSTEQIOFCFWKGFVVCVRGFEQKTRAPRPLMARLHPPASKLKNKT 234

RESULT 5
T05430
hypothetical protein F28A23.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C:Accession: T05430
R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.
submitted to the Protein Sequence Database, October 1998
A:Reference number: Z15415
A:Accession: T05430
A:Molecule type: DNA
A:Residues: 1-917 <BEV>
```

A:Cross-references: EMBL:AL021961
A:Experimental source: cultivar Columbia; BAC clone F28A23
C:Genetics:
A:Map position: 4
A:Introns: 41/3; 70/3; 88/1; 118/1; 321/1; 367/1; 474/1; 504/2; 624/3; 655/2; 667/3; 695
A:Note: F28A23.180
C:Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

Query Match 7.8%; Score 709; DB 2; Length 917;
Best Local Similarity 21.4%; Pred. No. 1.2e-30;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

QY 651 IAEIIRYQNLVGLGDKEREQONAMVLYKGDGALVPYESKRRKPRKVDIDETRIWNL 710
DB 388 IAKLIKMGRLKINK-----VTTMKADKLV-----TAKVNLDPETIKENDV 431

QY 711 LMGKDEGEDEEDKKEKWEERRVFRGRADSFIAHMLVOGDRFPWKGSVDVS 770
DB 432 LM--VNDSPRSYDDKETEAKWKEREIFQTRIDLFINRHLQGNRRKFKQWGSVDVS 489

QY 771 IGVFTQNVSHLSSAFMSLAARPPKLLSSRDERNVRVSVVDEPBGCTILNLEIPSW 830
DB 490 VGVFTQNTDYLSSNATMSVAKFP-----VDAREGLSYIEEPQ----- 530

QY 831 QEKVOHPDMEVSGVDYSGSKQELRDCNSGIERFNFLKESIQNLEEEVLSSQDSFDAIF 890
DB 531 -----DAKSSSECI----- 538

QY 891 QSCGRVSGSCSKSDAEPTTRCTKTVTSQSVQTSQSPNLSDIEICLOGNERPHLYEGS 950
DB 539 -----ILSDE----- 543

QY 951 GDVQKQETTNVAKKPDLEKTMNWKDSVCFGPRNDTNQTPSSSYEQCATRPHVLDI 1010
DB 544 ----- 543

QY 1011 EDFGMQGEGLGYSNWISPRVDRVKNKNVPRFRQGGVPREFTGQIIPSTHELPGMG 1070
DB 544 ----- 543

QY 1071 LSGSSAVOEHOQDQHNQOQDEMKNKASHLQKTFDLNLSSECLTRQSTKQNTDGLCP 1130
DB 544 ---SISKVEDHEN-----TAKRKNKKTGII- 565

QY 1131 RDRTAEDVVDPLSNNSLQNLIVESNSSNKEQTAVEYKETNATILREMKGLADGKPKTS 1190
DB 566 ---EDELVD----- 571

QY 1191 QWDLRKDVEGNEGRQERNKNMDSIDYAIRRASISEISAIKERGMNMLAVRIKDFL 1250
DB 572 -WNNLRR-MYTKEG--SRPEMHMSVNNSDVRLSGQNVLETTIKRQGFRLS----- 620

QY 1251 ERIVKHGIGDLEWLRESPDPKADKDYLLSIRGLGLKSVCEVRLTLNLNAPVDNTNGRI 1310
DB 621 ERIL-----VDNTNGRI 632

QY 1311 AVRGMVPLQPLPESLQLHLLLYPVLESIQKELWPLRCKLDQRTLYELHVQLITFGKVF 1370
DB 633 AVRGLVPLEPLPQVQMH-----QLFEYELHYOMITFGKVF 669

QY 1371 CTKSRPNACPMRGECHRFASAYASARLALPAPERSLTSATIPVPPEP-----PPVA 1425
DB 670 CTKTIPNCNACPMKSECKYFASAVSSKVLLESPEEK-----HEPNTFMAHSODVA 722

QY 1426 IPMTLEPLPLEKSLASGAPSNCEPIIEEPASPGOECTEITSIEDIAY-----NEDPD 1482
DB 723 VDMTSNINLVEECVSGGSDQAICYKPLVEFPSSPRAIPEST--DIEDVPFMNLYQSYA 780

QY 1483 EPTIKLNEIOFGMTLRHM--ERNMELOEGDMSKALVALHPHTTISPTP---KLKNLSR 1537
DB 781 SVPKIDFDLDALKSVEDALVISGRMSSSDEISKALVIPTPENACIPKIPPRKMKNYNR 840

QY 1538 LRTEHQVYELPDSHRLDGMKREPDDPSPYLLAIWTPGETANSAQPPEQKCGKGSKGM 1597
DB 841 LRTEHVYVLPDNLHDLDFERRKLDLDDPSVLLAIWQP----- 878

QY 1598 CFDETCSECNLSREANSQTVRGTLILPCRTAMRGSFPLNGTYFOVNLFAFHSSSLKPID 1657
DB 879 ----- 878

QY 1658 VPRDWINDLPRTVYFCTSVTSIFRGLSTBOIQCFWKGFCVCGVGFQKTRAPRLMARL 1717
DB 879 -----GFICLRAFRQKQKDPKELVRL 900

QY 1718 HFP 1720
DB 901 HTP 903

RESULT 6
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence:revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.1%; Score 278; DB 2; Length 5327;
Best Local Similarity 19.7%; Pred. No. 4.3e-06;
Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

QY 2 QSIMDSSAVNATEATEQDNGSRQDVLEFLNKTPOQPSKRKRKMPKVYVVG-----K 55
DB 1986 ESIDK-----EAKSKESSRE-----SVAESPLPSKEASRPASVAESTKDAEK 2031

QY 56 PKRPRKPAELPKYVVEGPKRKP-----RKAATQEKVKSKETGSAKKNLKESATKK 108
DB 2032 SKEESRRESVAESKPLPSKEASRPASVAESIKDEAKSKEESRRESAAESKPLPSKEASR 2091

QY 109 PANVG---DMSNKSPEVTLKSCRKALNFDLENPDGAR--OGDSE--SEIVONSSGANSF 160
DB 2092 PASVAESVKDEADKSKEESRRE-----SMAESGKAQSIKGDQSPLEKVSRPESVAESV 2144

QY 161 SEIRDAIGTNGSFLDSVSDIDKTNGLGMNQPLEVSMGNQPKLSTGAKLARDQDPDLL 220
DB 2145 KD--DPVKSKEPSRESVAGSVTADSDQSPLESKAGRPESVDSVDEAKQKESRR 2202

QY 221 TRNQOCCFPVATONTOPMENQQAWLQMLNQLIGFPFGNQOQPRMTIRNQOCLAMGNOQP 280
DB 2203 ESKTESVIPPKAOKDKSPKEVLQ-----PVSMTET--TIREDA-----DQP 2240

QY 281 MYLIGTRPALVSGNQ-----LGGPQGNKRPIFLNHQTCPLPAGNQLYGSTDMHQLV 333
DB 2241 M-----KPSQAESRRESIAESIKASSPRDEKSPSLASKEASRPGSVAESIKYDLDKPQII 2294

QY 334 MSTGGQOH---GLLKN-----QPGSLRIGQQPCVPLID--QQPATPK-----GF 374
DB 2295 KDKSTEHRSRESLEDKSAVTSKESVSRPLSVASDHEAAVAIEDDAKSSISPKDKSRPGF 2354

QY 375 THLNQVATSNSSPGLRPHSQSQSVPTTYLHVESYSR-----ILNGTTGTQCRSRAPAYD 428

QY 364 IDQPATPKGTHLNQMVATSMSPG---LRPHSQVPTTYLH--VESVSRILNLTG 417
Db 1562 LFOTPASGKDPVTVDETTKIALOSPOPHIINPASKMQSNMRLKDRFSSILEKOTQS 1621
QY 418 TCQSRAPAVDSLOODTHQNKYILSH-----ISNGCKK--ALPQNSSLPTPIMAK 469
Db 1622 RGRDAGTPA--PMQEE--NGTAIMETPKQKLDIFIGNSTGHKRRPTPKNRAOP-----1671
QY 470 LEEARGSKROYHRAMGOTEKHDLNLAQOI--AOSO-DVERHNSSTVEYLDAAKKTQIK 526
Db 1672 LEDLDGQELFOFPAGADSPVSVESAKISLASSQAEPVTPAST-----KRKSTGLSK 1726
QY 527 VVONLHGMPEVIEIEDDTGARKGKNTASISKGASKNSGSPVKKTABEKICVPIKTP 586
Db 1727 V-----DVRQEPSTLGRMK-----SLGRAPGTPAPVQ--ENDSTAFMETP 1766
QY 587 AKK-----GRGRKKSVPPAHASEIOLWQTPPKTPLSRSPKPKGRKSIQD-----SGK 637
Db 1767 KQKLDFTGSSGHR-----POTPKIRAP-----LEDLDGQEL 1802
QY 638 ARGPSGELLQDSIAELIYRMQNLGLDKEREQONAMVLYKGDGALVPYESKK--RKPR 695
Db 1803 FQTPAG--ANDSVT-----VEESVKMSLESSQAEPVKTPTAKRLSKTGL 1845
QY 696 PKVIDDITRINLLMGKDEKGEDEKKEKKEWEEERRVFRGRADSFAR-----749
Db 1846 SKVDREDPSILEKTKSPGTPAPVQENDCTAFMETPKQKLDFTGNSGHKRRPTPKI 1905
QY 750 ----MHLVQDGRFSPKGSVDSVIGVLTQNVSHLSSAPMSLAARPPKLSRSD 805
Db 1906 RAQPLEDLGQELFOFPAGASDV-----TVEESAKMSLESSQAEPVKTPTASTKRLS 1958
QY 806 ERNRSV-VVEDPEG---ILNLEIPSWQEKVQHPD-----MEVSGVDSGSK 850
Db 1959 KTGLSKVDVREDPSTGLKTKYKPCRAPGTPAPVQENDSTAFMETPKQKLDFAENS GSK 2018
QY 851 EQLRDCNSGIERNFLEKSIQNLLEEVLSQDSFDPAIFQSGRGVSGSCSKSDAEFFT 910
Db 2019 RRSRTSKN----RSOPLD-DLGDQELFOFPAGASNPVSVESAKI---SLESSQAEPVR 2070
QY 911 TRCETKTVSGT---SQSVONGSPNLDEICL-----QONE-----942
Db 2071 TRASTKRLSKTGLNKMDVRGHPGLSKSSCASQKVMQTLFLGDHGHRETQKGVLLAQKL 2130
QY 943 RPHLEGGSDVQKQETTNVAQKPKDL-----EKTMMNKDSVCFQPRNDTNWQTPSS 995
Db 2131 EPAIYVTRGRKQORCKRSQSPEDLSGVQEVFTSGHNKDSVT-----DNLAKLPSS 2184
QY 996 SYEOCATRQPHVLIDIEFGHOGGELGYSNWMSISPRVDRVKNKNVPRFRQGGSVPREFT 1055
Db 2185 S-----PPLETDTSVTSRRQARTGL-----RKVHVKN-----LSGGIMHPQIS 2224
QY 1056 GOIIPSPHELPGMG--LSGSSSAVQEHQDQTOHNOODENMKASHLOKTFDL-----1106
Db 2225 GEIV-DLPREPEGEKGVIKTKOSVKRKLDETVNVPKSKQRITRAKLTLEDLPQFQELC 2283
QY 1107 -----LNSSECLTRQSTKONITDGLCLPRDRTAEV---1138
Db 2284 QAPSLVMSDVIVKTPKMPDKSPPEVDTTSETQARRLRLLVVTPEIPORKTTRVVRQT 2343
QY 1139 ----VDPLSNSSQLNTL-----VESNSNKKQTA-1164
Db 2344 RNTQKEPISDNOGMEEPKSSVQKQDPSPVSLTGRNRQPTVKEKTPLEELTSFQEBETAK 2403
QY 1165 -----VYKETAATILREMGKTLA-DCKK--PTSOWDSLKDVEGNEGROERNKNN 1213
Db 2404 RISKSPQPEKETFLAGKKQLRIQLINDGQKEEPTAQ-----RKQ-PSRETRNLTKEPVG 2458
QY 1214 DSDIYEAIRRAISEISEAIKERGMNMLAVRIKDFLERIVKDHG-GIDLEWLR-----1266
Db 2459 DSINVEEVKSTKOKI-----DPVASVPVSKRRPRVPEKAQALELAGLQPIQTL 2509
QY 1267 ----ESPDK-----AKDYLLSIRGLGLKSVE 1289

Db 2510 GHYDESASDKGPTQMPCNLSLOPEQVDFSSPPRRPRTRRGKVEADEEPSAVR---KTVS 2565
QY 1290 CVRLLTLHLNLAFFVDVNG-----RIAVRMGW--VPLQPLPE 1324
Db 2566 TSR-QTMRSRKVPEIGNNGTQVSKASIKOTLDTVAKVGTSGRRQLRTHKGWSTLLKLLGD 2624
QY 1325 SLQI-----HLELYPVLESIOK-----FLWP-RLCKLDQRTLVHELHYQLITFGKVFC 1371
Db 2625 SKETIQISDHSEKLAHDTSLKSTQOQKPSVPLTCTCRVLRASKEVPKEVLVDTRDHA 2684
QY 1372 T-KSRPNACNPMRGECRHFASAYASARLAL-----PAPEERSLTSATIPVPPES 1420
Db 2685 TLOSKSNPLLSPPKRSARDGSIIVTRALRSLAPQEBATDEKVPPEKKRAASSRKYVSP- 2743
QY 1421 FPPVAIPMIELPLPLEKSLAS-----CAPSNRENCPIIEEPASPOECTEITESDIEDA 1475
Db 2744 --PVKMKHLKI-----VSNKLESVEEQVSTVMKTEEMEAARENPTVPO-----NS 2787
QY 1476 YYNEDPDEPTIKLNIQFQGMTLREHMERNMELQEGDMSKALVALHPTTTSIPTPKL---1532
Db 2788 RYRK-----KTNVKQ-----PRPKFDAS 2805
QY 1533 -----KNISRLRTEHQVVELPDSHRLDGMKREDDPSYLLAIWTGGETANSQAQP 1585
Db 2806 AENVGKKNKTKMTASQETEL-----QNPDD-----GAKKSTSR--2840
QY 1586 EQKCGKASKMCFDETCSECSNLSREANSQTVRGTLII---PCRTAMRGSP 1634
Db 2841 ----GOVSGK-----RTC-----RGTEMPOPCAEKTSKP 2871
RESULT 10
B84683
hypothetical protein At2g28300 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84683
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84683
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2218 <STO>
A:Cross-references: GB:AE002093; NID:g4803953; PIDN:AAD29825.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g28300
A:Map position: 2
Query Match 2.6%; Score 233.5; DB 2; Length 2218;
Best Local Similarity 19.1%; Pred. NO. 0.00031;
Matches 345; Conservative 252; Mismatches 644; Indels 563; Gaps 88;
QY 8 SAVNATEATEQDNGSRQDVLEF---DLNKTPQOKPSKRKRKFMKPVVVECKP--KRKPRK 62
Db 136 SAVSRQATGNAIGSSAATGLDFVSSDKRLEAASHPTSSALTSPTD--LSGPPGFQSLPAS 193
QY 63 PAELPVVVEGPKPKRPRKAATQEKVKSKETGS-----AKKNLKE 103
Db 194 PAPTP-IRGRGFRGRGAGRGRRVEGVVHGSSNTQRTTATATSLASDAEATKALPR 252
QY 104 SATK-----KPAWGVMSNKSPEVTLKSCRKALNFD--LENPGDARQG-DSES-----EI 150
Db 253 SASEIVSRVPKANEGSTSNPDQVSPVHSATTALRSDKAADKDLDPGFGDSGSHVQTLNV 312
QY 151 VQNSSGANSES-EIRDAIGG-----TNGSFLDSVSQDKTNGLGAMNQPLEVSGMGNPKDL 205
Db 313 LENSERKFAVKRPLIOGGGPVQNONAVSSVCDGSKSPSEGRTYTALO-GVTTAP---368

484	Db	AEKNIKNDILKPVQKSEGHKTKTKTQFETNNKQPGYNEKIMETGCKKINEDGCTRAQVEM	54
548	Qy	-----DGARKGNKTASISKGASKGNSSPVKKTAKEXKICVVPKTPAKKGRAGRKKSVPP	600
549	Qy		601
544	Db	IRQOELDEPARSEKENRSRELVSFKTNDSEKK---EKE---IAGTERKEKESDORPILRE	597
545	Qy		598
601	Qy	PAHASEIQLWQPPKTPPLTSRSPKPKGRKSIODSKARGPSCCELLCQDSABEIIIRMO-	659
598	Db	QEVADDEV-----AEDTKTSIYGEVKEEIEIAGKEFECSDDDDIARIVRDETEQ	645
660	Qy	---NLXLGDKREREQEQNMYLYK-----GDGALVPYPSK-----KRPRPKVIDDDE	703
661	Qy		704
646	Db	LDSNAMKQGEKQMIQELVEEKVCDGKGIIIAVAETKAENNKSKRVQETPEOKLDKEDT	705
704	Qy	TTRTWNLLMGKDEKDEKDEKDKKKEWEEERRV-PRGRADSFIAHRMLHVQGDRRFSPW	762
706	Db	CGKHQFKLIGETISDHGEVEDVEKGGKRTAEKRKIDKRAEAEIEKDLGVSGRYI---	762
763	Qy	KGSVYDSVI---GVFLTONYSHDLSSSAFNSLAARFPFKLSSREDBERNVRSVVVEDPEG	819
764	Qy		820
763	Db	KGTTIKELVENRGIYRNEH-----EEXKKDDAN-----RPEK	794
820	Qy	CILNLNEIPSWOEKVQHPSDMEYSGVDGSGK-----EOLRDC-SNSGIE-----RNFLE	868
795	Db	ITGTIK-----QELVLSNLSQRLQENVEDGDKTOELVEEKIKDCEEEGSESKIKTDVVV	849
869	Qy	KSIONLEEEVL-----SSQDSFDPALFQSCGRVSGSCSK	903
870	Qy		904
850	Db	RKVQIGKEEELYKPKREHGTKITELVEETTGVDYEQKEKETAEASDIEAECG-----SLRK	904
904	Qy	SD-----AEPFTR-----CETKTVSGTSSQSVQTGSPN	931
905	Db	VDTIEEHHELHPKTHKDRNNRVYTGAKPEQSGQKEGKEEKIVESMTITENDNSIDVQETK	964

Q1 965 KERPGRLESHDKRYKIQELLMEAGHNDKREEQENNVTAVELETETRVSSKKVQEGMED 1024

[illegible]

RESULT 13
T20513
hypothetical protein F02E9.4 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #t
C;Accession: T20513

R.Gray, I.
submitted to the EMBL Data Library, November 1996

A:Reference number: Z19285

A:Accession: T20513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1490 <WIL>

A:Cross-references: EMBL:Z81494; PIDN:CAB04052.1; GSPDB:GN00019; CESP:F02E9.4

A:Experimental source: clone F02E9

C:Genetics:

A:Gene: CESP:F02E9.4

A:Map position: 1

A:Introns: 225/3; 321/2; 373/3; 474/3; 619/3; 664/3; 697/3; 828/3; 868/3; 902/2; 965/2;

Query Match 2.4%; Score 217.5; DB 2; Length 1490;

Best Local Similarity 19.2%; Pred. No. 0.0013;

Matches 315; Conservative 221; Mismatches 555; Indels 547; Gaps 86;

Qy 281 MYLIGTPRALVSGNQQLGGPGGNRP-----IFL-----NHOTCLPAGNQLYG 324

Db 1 MY---NPPPGGGGNN--GGDSQQOQPTNNATLFLQMIQQSQHQOQHQN--QQQQQLEL 53

Qy 325 SPTDMHOLVMSTGGQGHLLIKNQPGSLIRGOQPCVPLIDQOPATPKGFTHLNQTWATS 384

Db 54 QIRDERILIEQRMQH-----QQQONLQO-----LNQFPNPLGLFQVQAQAA 101

Qy 385 MSSPGLRPHSQ--SQVPTYLHVESVSRILNLTGTCQSRAPAYDSLOQDIHQNKYILS 443

Db 102 QAO--FAQAQGSPIP--FHIGS-----PLQSHSPAASALQ-----QYL-- 138

Qy 444 HEISNGCKKALPONSLPTPIMAKLEARGSKR-----QYHRAMGOTEKHDLNLA 495

Db 139 -----LPSHSPAITPFPARNSEARNIEFIAQEEAANVPRANSQQS---PLI 183

Qy 496 QQIAQSQDVERHN---SSTCEVYLDAAKTKIQVQENLHGMPEVIEIDDPDGTARK 552

Db 184 RPIPOQALNTQNLSTQAAQILAHQVQVQVQHQHQTPTPLALPI-----AQO 236

Qy 553 GKNTASISKGASKGNSPVKTAKEKICVPTPA-----KKGRAGRKSV 598

Db 237 G-----PISNEVPVPVPVVPATSGCPQREPQOQGGRRQNRPKRRKP 280

Qy 599 PPPAHASELIQWOPPTKPTPLSRKPKGKRSKIDSKARGPSGELCQDSIAEIIYRM 658

Db 281 EGPRVDEALAYLRVIKSTFSSDVPVYHREIMKDFRAQRIETPDVI--EQVALLYDS 338

Qy 659 QNLYLG-----DKE-----REQQNAWLYKGDGALVPYES 689

Db 339 PELVLGFNTFLPTGYRIITLTPDRKIVFSSPOMOPRVLLSPDERRARAIEAQAQVGAIEL 398

Qy 690 KKRPRPKVIDDETTRIWNLLMGDEK-----EGDEE---KDKKKEKWEEERRV 738

Db 399 GSQEGISK---DEDRTIEDEMDKSEKDDVGDGIDDEDESGIEDKNNEEMEDNHL 454

Qy 739 PRGRADSIARHLVQDRRSPKRGSVVDSVIGVFLTQNVSDH-----LSSAFMSLA 792

Db 455 IE-----EITCDRK---KDDCEDSQOQETMSSELAHAHTLNIIELLKKSFL--- 497

Qy 793 ARFPKPL-----SSSREDERNVRVSVVDEPEGI-----LNLNEIP---SWQEKV 834

Db 498 AR-PTKLVDFMTFIDFFMSDQYKKDMKELKDDDEDEIEENEKIEVDVDPGSPNAPQEI 556

Qy 835 QHPDMEYSGVDSGSKQLRD--CSNSGIERFNFLKSIQNL---EEEVLSSQDSFDPDAIF 890

Db 557 KPDDIEKK--DSSKNLQIESCS-----DYLVSMLANCCIGEPDLAATIDPLPYL- 606

Qy 891 QSCGRVSCSCSKSDAEPTTRCTETKTVSGTSQSVQTSQPNLSDBICIQGNRPHLYEGS 950

Db 607 -----GKLLVNGSDA-----TALKIKTILHFSAT-----NDR----- 633

Qy 951 GDVQOQETTNVAQKKPDLEKTMNWKDSVCFQGPQNDTNNWOTTPSSSYEQC--ATRPQPHVL 1008

Db 1008

RESULT 14

T48818

glucan 1,4-alpha-glucosidase related protein [Imported] - Neurospora crassa

N:Alternate names: protein 68B2.20

C:Species: Neurospora crassa

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000

C:Accession: T48818

R:Schulze, U.; Align, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, April 2000

A:Reference number: 224541

A:Accession: T48818

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2022 <SCH>

A; Cross-references: EMBL:AL353821; GSPDB:GNO01112; NCSP:68B2.20
A; Experimental source: cosmid contig 68B2; strain 74
C; Genetics:
A; Gene: NCSP:68B2.20
A; Map position: 2
A; Introns: 1192/3

Query Match	2.4%;	Score 217;	DB 2;	Length 2022;
Best Local Similarity	17.2%;	Pred. No. 0.0022;		
Matches 340:	Conservative 253;	Mismatches 720;	Indels 668;	Gaps 85;

Qy	35	PQOKPSKRRKRMPKVVBEGPKRPRKPAELPkvvVEGKPKRKPRKA-----ATQEKKV	89
Dd	4	PKPKDK-----PSPRPKTFTPTNSKNKGAPTISTKSATAGATTPKSV	50
Qy	90	SKETSAKKNNLKESATKK--PANVGD--MSNKSP-----EVTLSCKRKALNFLENPGD	140
Dd	51	PTPTNSGLSHHKKPELKKPEPSLLGDFLGRPSQPORVAQAQRASAKRRTKSMDAQNVRE	110
Qy	141	AROCDSESEIV---QNSSGANSEFTRDAIGTNGSFLDSVSQIDKTNGL-----GAMN	191
Dd	111	ELROEMRAAAVRKLQOPGG-----VDRVKAWQAKASQNAV---KAGELPIPVAEADARS	160
Qy	192	OPLVISM--GNOPD-----KLSTGAKLARDOQPDLLTRNQOCQFPVATONTQFFMENQQ	243
Dd	161	EPTFAVLNSDDDEEDRMRIKWQKPPKQKKKIET-----VEPKGT-----SGSD	208
Qy	244	AWLOMKNLGIFFPEGNOQPMRTTRNQOCPCLANGNOQPMYTLIGHTPRPALVSGNQOLGGPQG	303
Dd	209	STAKEKNKVLG-RMADDPDVIIKEARP-----RP-----GLPKRIISLDHWMKKQO-	255
Qy	304	NKRFIFLNHTQLPAGNOLVGSPTDMHQLVMSTMGGOOHLIKNOOPGSLIRGOQCVPVL	363
Dd	256	-----KGKGLAKSP-----VPEAQO-----PIPKDFLARTA-----	281
Qy	364	IDQOPATPKGFTHLNQMVATSMSPGLRPHOSQOVPTTYLHVESVSRILUNGTTGTCQRSR	423
Dd	282	--QNPSVOKKIRDAWAORVELPPPPP--PVARRPTVKTRYHART-----GETVTFVEDE	330
Qy	424	APAYDSLQODIHOGNKYILLSHEISNGCGCKKALPONSLPTPTIMAKLEARGSKRYIHR	483
Dd	331	-----DDDHHRG-----ALSHPD-----PRKRSS	351
Qy	484	MQOTEKHDLNLAQIAQSOD-----VERHNSSTCWEYLDAAKTKIQKVQENLHGMP	537
Dd	352	SGSERSSSNPRTQCCPNDBGIRVRSLOKRGSD-----DGIRISVPKPA-----RSLPD	401
Qy	538	EVIEIEDDP---TDGARKGNTASIS-----KGAS-----KGNSSPVKK	573
Dd	402	DGIVRFGPPVYSADSSSRKSISTVSPSSSGORTPSDRSGSTPPRGASPPPRASTPPRRA	461
Qy	574	TAEKECVTPKTAAPKAGRAGKKSIV-----PPPAHASEIQLWMOPTP	615
Dd	462	STPLPKASTPKPARSDHSASDDVIEIVPEPSEVSKRSPSPPKR---RLRSPPPP	517
Qy	616	KTPLRSRK-----PKGKRKSIOD	634
Dd	518	KRKLPRRRGSGGARRKPRRSPSPPTTATQTETTTDDRRPGADKPMTPRNNGSGSGED	577
Qy	635	SGKARGPSGELLCOdstAEIIY---RMQNLVLDKEREQONAMVLYKGDGALVPYESKK	691
Dd	578	SDR-RPPTAAAGI--DDUAEIPFGFSAFSELPLURGHUTQTHA-----ROS	620
Qy	692	RKPREKVIDDETTRIWNLL-----MGKDEKEGDEEKDKKEKWEERAV	738
Dd	621	AKPREKPORNESKLVPNLVKVMTGAEKMOEMAEP RPPTGTGNKPARTES- - - - -	673
Qy	739	FRGRADSTIARMHLVQGDREPSWKGVSDSVIGVFILTQNVSDHLSSAFMSLAAREFPK	798
Dd	674	LNNTVDPFEVGMNPNI.PPVVAPEPLRVSTPE-----RNSKEKLVDRDLP.AHRERAPER	725
Qy	799	LSSSREDERNVRSVVVEDPEGCILNINEIFSWOEKVQHP-----SDMEVSGV	845

Db	726	NSKE	LDVTDVSTHKARTPEISKRASAELEPRRK	POTPVYTHEELKRTSTRSDSEVTPK	785
Qy	846	DSGS	KEQLRDSNSGIER-----	FNFLEKSIQNLEEVL-----	SQSDSDPAI 889
Db	786	RRKS	PTTTPD-SAAGLKRKSTPRSTPLPRSSGSVKKPF	RDLVLEAKFGESSAHKIAPMV	844
Qy	890	FQSC	RGVSCSKSDAEFPITR	CTETKVTSCTSQSVOTGSPNLSDEICLQGNRPHLYEG	949
Db	845	YPSC	-ETDVESEPTEHLESRRSPPOORS	PDSTKYKRASAP-----	DRPSRAD 892
Qy	950	SGDV	KOETTINVAOKP-DLE--	KTMNKSVCFGQP--	RNDTNMQTTPSSSYEOCATR 1003
Db	893	YSSP	ESTRGPSSRRRPTSDLHLSL	ILSEDSREYQEPYCKDITSVTSQS--	TVTQ 949
Qy	1004	QPHV	LIEDFGMOGELGYSWMSIS	PRVDRVKNKNVPRFRFGG	SVPRET----- 1055
Db	950	APE	---EAFGP-----	TPUTREASQOSQTSIRKSGSLKRLTKHSDLSV	994
Qy	1056	---	GQIISTP-----	HELPGMGLSGSSAVQEHQDDTOHNOQDEMKNASH	1098
Db	995	VLSL	PDGGLVPPSSRSIKASRSLHRKPS	KANDSRVNDLLEEFADDEHYHREL-----	1049
Qy	1099	LQKT	FLD-----	LLNSBECLTRQSTKQNTDGC	LPDRTAEDVD----- 1140
Db	1050	--KTL	VDGVVPLLN-----	EFVHGDNVDDADRTDSMAKAVNMVGALEK	LWYTHK 1099
Qy	1141	---	-----PLSN-----	-----	NSLQNILVESN 1156
Db	1100	RAPL	DIRLLEWLEAVSPVYNNYLDVWRL	FGDILVNLAPPSKIDENDSLNALPRNE	1159
Qy	1157	---SS	KEQTAVEY-----	---KETNATILREMKGT	LADGKKP 1188
Db	1160	DGVL	YNGERVDVAYLLKRP	LIRTKWYKFLRAVWIKTPETEDLLKLYGNLOEKART	1219
Qy	1189	TSOW	SLR-----	KDVEGNEGRQRNNMDSIDYEAIRRASTISEA	IKERGMNML 1242
Db	1220	RHRE	EARMTDEDANNTDTRARDL	RNLRLPDLNVRIDPSQVAAKSFEMDLHSSGQRL	1279
Qy	1243	AVIK	-----	DFLERIVKOHGGIDLEWLRSP-----	---PKAKDYLLSIR 1281
Db	1280	ECQV	ELMRDRVNPVSDMGDILIRDIS	KNACKPWLFPVPVPROYISARKGSPRSIMVIR	1339
Qy	1282	GL--	GLKSEVCVRLTLHLNLA	FPVDTVNGRIAVRMGWVLPQLP-----	PESLQL-HLEEL 1333
Db	1340	GRING	DWYELIKLSTEEVQITDLGI	-----	LCSDMPPLSRPXPMSLIHVSS- 1392
Qy	1334	YPVLE	STQKFLPRLCKDQRTLYELHY	QITFGKVCTKSRPCNACPMRGECRHFASA	1393
Db	1393	-----	PKADDL-----	VPIGEPSVIGVDDEFRTPSYHKKQASAP	1429
Qy	1394	YASAR	LALAPERSITSATIPV--	PPESFPVPAIPMLPLPLE-----	KSILAS 1441
Db	1430	VTS---	YPAPNKHIPYSVDIFRSQITPA	YGIESVIPRPLNIKGSVAVQPVELAP	1485
Qy	1442	GAPSN	CEPILEE--	PASPGOECEITEIDEDAYNED-----	PDEIPTIKLNE 1492
Db	1486	TRVR	KRTSSPLKHEXHP	SDISETEESARSDESSSSSDELDDEYDPTIP-----	1538
Qy	1493	QFGM	TLREHMERNNELQGDMSKAL	VALHPHTTISI-----	PTPKLNISR----- 1537
Db	1539	--GYS	LAKPLQIIVE--	ESVVSENTIAPSQASQVGAPOPGOOT	POERTVQRFVASVYW 1594
Qy	1538	-----	LRTE-----	---HOVELPDSHRLD-----	---GMDKREPDDPS 1566
Db	1595	SNRKG	IWEINTEPSRILVPPGSM	YHMLQETPGNKQAYPLQTS	GVSEVDMANREAGV 1654
Qy	1567	PYLLA	IWTP-----	GETA-----	NSAQPPQKCGGKAGKMGCFDETC- 1615
Db	1655	PLIC	LVITPVVMIRRTALDLEVR	SRVSPESRLSSIDSQMFPRASODEAKALYEAHVQ	1714
Qy	1616	T	1616		

Db 1715 S 1715

RESULT 15
T05324
hypothetical protein FIC12.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05324
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15408
A:Molecule type: DNA
A:Accession: T05324
A:Residues: 1-1188 <BEV>
A:Cross-references: EMBL:AL022224
A:Experimental source: cultivar Columbia; BAC clone FIC12
C:Genetics:
A:Map position: 4
A:Introns: 305/3; 348/2; 975/2; 1091/3; 1141/1
A:Note: FIC12.80

Query Match 2.3%; Score 211; DB 2; Length 1188;
Best Local Similarity 17.4%; Pred. No. 0.0021;
Matches 244; Conservative 219; Mismatches 532; Indels 406; Gaps 55;

QY 78 KPKAAATQEVKSKTSGSAKKNLKESATKKPANVGDMNKSPEVTLKSCRKALNF----- 133
DB 20 RPERCTQ---RDNTNNKKNN-----NNNNVRNIHVAEKNLNVEVRD 63

QY 134 DLENPG---DARQDSSEIVQNSSGANSFSEIRDAIGGTNGSFLDSVSQIDKTINGLGM 190
DB 64 HLENCVSVDVDDVDSIKAVPECSNKSNSVSDHR--LSKTEESRQDVPS---SSSISLGS 118

QY 191 NOPLEVSNQPKLSTGAKLARDQDPLTRNOQCFPVATONTQPMENQQAOLQMN 250
DB 119 SDP-----NSGQPNPSRNAASLVQIWEART-----TOQPPSSNQLSDSRT 161

QY 251 QLIGF-PFGNOQPMRTIRNO---OPCLAMGNOQPMYLIGTPRALVSGNOQLGGPOGNKR 306
DB 162 SSMGSNALENSSESLESVKSQMIQPIECNNEEBEEIECAPPLESGEKDREGVRVMDI 221

QY 307 PIFLNHTCLPAGNOLYGSPTDMHQLVMSTGGQQHGLLIKNNQPGSLIRGOQPCVPLIDQ 366
DB 222 IRKLSNDSSETITNDNGSSGNDNSKEVQTTEARSPQVACSPR---IRGRQALADLLVQ 277

QY 367 QPA-TPKGFTHLQNVATSMSSPLRPHSQSOVPTTTLHVSVSRILNGTTGTQCRSRAP 425
DB 278 MTRDREKDLACLREHRCVSKFT-----NRGRIQST-LRIRCVCYCLAIQGRHRSKSTSA 330

QY 426 AYDSLOQDIHQGNKYILSHETSNGCKKALPONSLPTPIMAKLEARGSKROYHRAWG 485
DB 331 GSDNRSRSGVMHLLREKY-----KANSENIETGASTSHASTGRIMDKDPHKA-- 381

QY 486 OTEKHLNLAQOIAQSODVHRNNSCTVEYLDAAKTK--IQKYVOENLHGMPP---EVI 540
DB 382 -TEKKVL-----QETIEKSGVKKEIKVVSAYENAKKGVLSSEISNRNGLKLAGETI 432

QY 541 EIEDDPTDGARKKNTASISKAGSKGNSPVKTAEEKIVPKTPAKKGRAGRKKSVP 600
DB 433 QKETVEGKGEKRETKKRVISKESLEGKE--KRESTKEKAIAKESVAEKALVG----- 483

QY 601 PAHASEIQLQOPTPKPTPLSRSPKGGKSIODSGKARG----- 640
DB 484 --IAEKNLW-----NSDEKNMRKRVKTEGNTNTINTERVNDVLEEFATPRI 530

QY 641 -----PSGELLQODSIAETIYRMQN-----LYLGDKEREQENAMVLYKGDGA 683
DB 531 LSVESAERSSTTTSKETMTCEVAEKVVVKKKEDFVLSLESNKSKEVEDGNI---KPGQV 587

QY 684 LVPYESKKRPRPKVIDDETTTRINLLMGKGDEKGEDEKDKKKKKEWEEERRVFRGRA 743

Db 588 TQADSLCRKPAIEEKVLQETT-----VKSDLKKPTEKSEEREEREIEEESTSMGIA 640

QY 744 DSFIARMHLVQDGRREFSPWKGSSVDVSVIGVFLTONVSDHLSSSAFMSLAARFPKLLSSR 803

Db 641 EKVNLNMSKEKKNRKAMEKKG--GKTEGKAITE-----TNELQAEAR---RISNVE 688

QY 804 EDERNVRSVVVEDEPGCILNLEIPSMQEKVQHPSDMEVSGVDSGSGKEQLRDCSNSGIER 863

Db 689 TAERSIDT-----SWITVVVGVDRVINDKRKSSGETTSAEIGGKK 731

QY 864 FNFLEKSIONLEEEVLSSQSDFPDPAIQSCRGVSGSCSKSDAEFFPTRCTETKTVSQTQ 923

Db 732 EEDL-ASVEAKSKDVIDEKNMNPQAVIH-----GS-----KERDKE-----RN 768

QY 924 SVGTGSPNLSDCLCLOGNERPHLYEGSDVQKQETT-----NVAQKKPDLEKTMNWKD 976

Db 769 SSQNGEETLS-----LRNSEA-----KSTKEIEREVTQOEKSVSHGSKREKKNLSQYGE 820

QY 977 SVCFGQPRNDTNMQTTPSSSYEQCATRQPHVLDIEDFGMOGEGLYGYSWMSISPRVDRKN 1036

Db 821 KMCF--LRNS-----EAKSTKE-----IERNKS 841

QY 1037 KNVPRREFROGGSVPREFTGQIIPSTPHELPGMGLSSSSSAVQEHQDDTTOHNOQDENKA 1096

Db 842 QEV-----SQGEESASHGSRSAKEKNSQODD----- 869

QY 1097 SHLQKTFDLNLSSECLTQROSSTKQNTDGCCLPRDRTAEDVVDPLSNSSQLNVLVESN 1156

Db 870 -----ETSTHRN-----PNDK--KGIKEPEDE-----ESK 892

QY 1157 SSNKEQTAVYEKTNATILREMGKT-----LADGKKPTSQWDSL 1196

Db 893 KVEREETGENVEEASVEFVNDWDGNEEMEEEEEYGDYFNGDDDDWIHDISRPSYWEDLR 952

QY 1197 KDVEGNEGROERKNKNMDSIDYEAIRASISEISEAIKERGMNMLAVRIKDFLERI--- 1253

Db 953 KE-----RYLEVLNTESEKK---DICNLIERRTVSNFLTSLRQKIDNLMIT 996

QY 1254 -VKDHGGIDLEWLRESPPDKADYLLSIRGLGLKSVBCVRLTLHLNLAFFVDT---NVGR 1309

Db 997 RVOSHIGVPLNQIEEG-DEYEEWEVECSARNEDNETEEPEKTNLEAPSDVCSQSSAR 1055

QY 1310 IAVRMGW-----VPLQPLPESLQ--LHLELYPVLESIQKFLWMPR 1347

Db 1056 SSTMMSNFRDQDIDKDNPTTSLSLPELVPTNQSTQMDQTTISDLKEOMEQLOREM--- 1112

QY 1348 LCKLDQRTL-----YELHYQ 1362

Db 1113 ---LELRNTVKSCIDMQLHFQ 1130

Search completed: September 27, 2002, 12:57:18
Job time: 9052 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:55:52 ; Search time 33.36 Seconds
(without alignments)
2006.778 Million cell updates/sec

Title: US-09-840-743-2
Perfect score: 9089
Sequence: 1 MQSMDSSAVNATEQND.....PRPLMARLHFPASKLKNKNT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	246.5	2.7	3256	1	KI67_HUMAN
2	225.5	2.5	2492	1	ATRX_HUMAN
3	225	2.5	2468	1	MAPB_HUMAN
4	219.5	2.4	5430	1	ACFT_HUMAN
5	211.5	2.3	2842	1	APC_RAT
6	210.5	2.3	1805	1	NEST_RAT
7	208.5	2.3	3924	1	ANK2_HUMAN
8	207	2.3	1723	1	AIM1_HUMAN
9	204.5	2.2	2717	1	ZEPI_HUMAN
10	200.5	2.2	3130	1	DPO2_HUMAN
11	192.5	2.1	1781	1	AKAC_HUMAN
12	192.5	2.1	1812	1	BRCL_MOUSE
13	192	2.1	2442	1	CBP_HUMAN
14	191.5	2.1	2349	1	TPR_HUMAN
15	190	2.1	1284	1	ATL_COMPX
16	190	2.1	2464	1	MAPB_MOUSE
17	189	2.1	2453	1	NCRI_MOUSE
18	188.5	2.1	1435	1	EBAL_PLAFC
19	188.5	2.1	3396	1	PGCV_HUMAN
20	188	2.1	2805	1	MAPA_HUMAN
21	187	2.1	3911	1	AKA9_HUMAN
22	186.5	2.1	1132	1	YKKS_YEAST
23	186.5	2.1	1833	1	ZEPI2_HUMAN
24	186	2.0	1531	1	NFT5_HUMAN
25	184.5	2.0	1875	1	MLP1_YEAST
26	183	2.0	2441	1	CBP_MOUSE
27	182.5	2.0	1861	1	MAP2_RAT
28	182	2.0	2688	1	ZEPI_MOUSE
29	181.5	2.0	3381	1	PGCV_BOVIN
30	181	2.0	1658	1	YM67_YEAST
31	180	2.0	3122	1	DPO2_MOUSE
32	178	2.0	2230	1	GOGA_HUMAN
33	178	2.0	2476	1	ATRX_MOUSE

ALIGNMENTS

```

RESULT 1
KI67_HUMAN
ID KI67_HUMAN STANDARD; PRT; 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen KI-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-94043435; PubMed-8227122;
RA Schluter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CHROMOSOMES.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS
CC PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL
CC CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.
CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X65550; CAA46519.1; -
CC EMBL; X65551; CAA46520.1; -
CC EMBL; X94762; CAA64388.1; -
CC MIM; 176741; -
CC InterPro: IPR000253; FHA_domain.
CC Pfam; PF00498; FHA; 1.
CC SMART; SM00240; FHA; 1.
CC PROSITE; PS50006; FHA_DOMAIN; 1.

```

Q12888 homo sapien
Q62059 mus musculu
Q03001 homo sapien
O61315 mus musculu
P51825 homo sapien
O14497 homo sapien
Q13061 homo sapien
P10388 triticum ae
P49454 homo sapien
P08799 dictyosteli
Q09472 homo sapien
P55200 mus musculu

Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;

KW	Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;	588	KKGR----	AGRKSVPPPAHASE-----	IQLWQTP-----	PKTPLSR	621
KW	Alternative splicing.	27	76	FHA	16 X 122 AA APPROXIMATE REPEATS.		
FT	DOMAIN	1000	2928	1. 1112	1. 1112		
FT	REPEAT	1000	1112	2. 1234	2. 1234		
FT	REPEAT	1222	1234	3. 1244	3. 1244		
FT	REPEAT	1244	1356	4. 1477	4. 1477		
FT	REPEAT	1366	1477	5. 1598	5. 1598		
FT	REPEAT	1487	1598	6. 1720	6. 1720		
FT	REPEAT	1608	1720	7. 1842	7. 1842		
FT	REPEAT	1730	1842	8. 1964	8. 1964		
FT	REPEAT	1851	1964	9. 2086	9. 2086		
FT	REPEAT	1974	2086	10. 2204	10. 2204		
FT	REPEAT	2096	2204	11. 2326	11. 2326		
FT	REPEAT	2214	2326	12. 2447	12. 2447		
FT	REPEAT	2335	2447	13. 2569	13. 2569		
FT	REPEAT	2457	2569	14. 2689	14. 2689		
FT	REPEAT	2579	2689	15. 2808	15. 2808		
FT	REPEAT	2699	2808	16. 2928	16. 2928		
FT	REPEAT	2818	2928	ATP (POTENTIAL).			
FT	NP_BIND	3034	3041	MISSING (IN SHORT ISIFORM).			
FT	VARSPLIC	136	495				
SQ	SEQUENCE	3256	AA: 358741 MW: 578F8C51BED42517 CRC64;				
Query Match 2.7%; Score 246.5; DB 1; Length 3256;							
Best Local Similarity 18.8%; Pred. No. 7.9e-05;							
Matches 362; Conservative 266; Mismatches 768; Indels 527; Gaps 94;							
QY	22	SQDVLEFDLNKTPQOKPSKRRKFKMPVVVEGK----	PKRKR--	KPABELPKVVVEGKP	75		
Db	1496	SQDPVDPVTPSSKPSKRRSLR-----	KVDEEFFALRRKTRPSAGKAMHTPKPAVSGEK	1549			
QY	76	KRPKRAATOKYKSETGSAKKNL-----	KESA-----	TKKPAVGDMSNK	118		
Db	1550	NIYAFMGTPVQKLDLTENLGRSRRRLQTPKEKAQALDLAGFKELFOTRGHTESMTNDK					
QY	119	SPEVTLKSCRKALNFDLPDNGADQDSESEIVQNSSGANSFSFEIRDAIG-----	GTNG--	172			
Db	1610	TAKVACKSQPDLD--	KNPASSKRR-----	LKTSLGKVGKVEELLAVGKLTQTSGETT	1660		
QY	173	-----	SFLDSVSI--	DKTNGLGAMNQPLEVSMNQ--	PDKLSTGAKLARQD	215	
Db	1661	HTHTEPTGDKSKMAFESPKLIDSAASLTGSKRQLRTPKGSSEVPEDLAGFELF--	Q	1718			
QY	216	QPDLLTRNOCCQPPVATONTQFPMENOQ----	AWLQMKNOLIGFPFGNO--	QPRWTIR--	267		
Db	1719	TP-----	SHTKESMTNEKTKVYRASQPDLDVTPTPSSKPPQKRSURKAD	1763			
QY	268	NQOPLCANGNOQPM--	YLIGTPRALVSGNQQLGGPQGNKRPIFLNHTCLPAGNOLYGS	325			
Db	1764	TEEEFLAFKQTPSAGKAMHTPKPAV-----	GEEKDI-----	NTFLGT	1801		
QY	326	PTDHHQLVMSTGGQOCHLLIKNOQPSL--	IRG-----	QOPCV--	PLIDQOPATPKGFTH	376	
Db	1802	PVQKLDQPNLPSNRRLQTPKEKAQALELTGFRELFTQPTCDNPTADEK-----	TT	1854			
QY	377	LNQWATSMSSPGLRPHSOSQVPTVLHVESVS-----	RILNGTGTCCORSRAPVDSL	430			
Db	1855	KTILCKSPQSDPADTPTNTKQPKRSLLKADVEEFLAFKLTSPSAGKAMHTPKAAVGE--	1913				
QY	431	QODIHO--	GKNYILSHSINGNGCKKALPONSLPTPIMAKLEEARGSKROY-----	HRAM	484		
Db	1914	EKDINTFVGPVPEKLDLLGNLPSKRR--	PQT-----	PREKAKALEDLAGFKELFOTPGHTEE	1969		
QY	485	GQTEKHDLNLAQIAQSQDVVERINSSCTVEYLDAAKTKIKQVQVENLHGMPEVIEIED	544				
Db	1970	SMTDDKITEVSCSKSPQDPVKTPTSS-----	KORLKISLGKV-----	GVKEEVLVP--	2015		
QY	545	DPTDGARKGNKTASISKGA-----	SKGNSSPVKKTAKEKCIV-----	PKTPA	587		
Db	2016	-----	GKLTQTSKTTQTHRETAGDGKSIKAFKESAKQMLDPANYGTGMERNRPTPK	2067			
QY	588	KKGR----	AGRKSVPPPAHASE-----	IQLWQTP-----	PKTPLSR	621	
Db	2068	EAAQSLEDLAGFKELFOTPDHTEESTDDTKTKIACKSPPEESMDTPTSTRRPKPLGK	2127				
QY	622	-----	SKPKG--	KGRKSTODSGK-----	ARGPSGE	644	
Db	2128	RDIVEELSALKQLTQTHTDKVPGDEDKGINVFRETAQKQKLDPAASVTGSKRQPRTEKKG	2187				
QY	645	LLQODSTAEIIRYMONLYLGDKEREQEQANVLYKGD-----	GALVPYESKKRKRPRKVD	699			
Db	2188	AQPLEDLAGLKFOTPVCTDKPTTHEKTKTIACRSPPQDPVGTPTTIFKPSKRSRLKAD	2247				
QY	700	IDDETTRIWNLLMGKDEKE-----	GDEEKDKKKWEEERRVFRGRADSFARHML--	752			
Db	2248	VEEESLALRKRTPSVGKAMDTPKPAGDEKDMK-----	AFMGTP--	VQKLDLPG	2294		
QY	753	VOGDRRFSPW-----	KGSVDSVIG--	VFLTONVSHLSSAFMSLAARPP-----	796		
Db	2295	NLPGSKR--	WPQTPKEKAQALDLAGFKELFOTPGTDKPTTDEKTKIACKSPQDPD	2351			
QY	797	PKLSSSREDERNVRSVVVEDPEGCILNLEIPSWQEKVQHPSDMEVSGVDSGSKQLRDC	856				
Db	2352	TPASTKORPKRNLKADVE--	EFLALRKRTPSAGKAMDTP-----	KPAVSDE	2397		
QY	857	SNSGIERFNFLEKSQNLNLE-----	EVLSSQDSFDPAIFOSCG-----	894			
Db	2398	KNIN----	TFVETPVQKLDLLGNLPGSKRQPOTPKEKALEDLVGF--	KELFOTPGHTEE	2452		
QY	895	-----	RVGSCSCSKDAE--	PPTRCE-----	TKTVSGTSQSQTG	928	
Db	2453	SMTDDKITEVSCSKSPOPEPKTSRSKQRLKIPLVKVDKKEEPLAVSKLRTSGETTQTH	2512				
QY	929	SPNLSDIEICLOG--	NEHP--	HLYESGDV--	QOETTNNVQAKKPDLEKTMNKWSDVCQOPR	984	
Db	2513	TETGSKSIKAKESPKQLDPAASVTGSKRQLRTRKARALELDVDFKE--	LFSAPG	2570			
QY	985	NDTNWQTPSSSYEOCATRQPHVLDIEDFGMOGEGLYGYSWMSISPRVDRVKNKNVPRRF	1044				
Db	2571	HTESMTIDKNTKIPCKSPPELTD-----	ATSTKRCPKTR--	PRKEV	2612		
QY	1045	ROGGSVPRETGQIIIST--	PHELPGMGLSSSSAVQE-----	HQDDTOHNOQDMMK	1095		
Db	2613	KELSAVERLTQTSOSTHTHKEPASGDEGIKVLQKRAKKKPNPVEEPPRRRAPKKEK	2672				
QY	1096	ASHLOKT--	FLDLLNS--	EECLTRQSSQNTDGLPRDRTAEDVDVPLNSSLQN	1150		
Db	2673	AQPLEDLAGFTELTSESGHQSLSLTKATK-----	IPCESPPLEVDVTTASTKRHLR	2725			
QY	1151	ILVESNSSKEQTAVEYKEFNATI-----	LREMGKTLADGKKPTSQMDSLKRKDEGNE	1203			
Db	2726	TRVOKVQVKEEPSAVKFTQTSGETTDADKEPAGEDKGKA--	LKESAKQTPAPAASVTGSR	2784			
QY	1204	GRERKNKNWDSIDYAIRRASISEISEAIKERGMNMLAVRIKDFLERIVKHGGIDLE	1263				
Db	2785	RRPRAPRESAQATE-----	DLAGEKDPAAAGHTEESMTDDTKTKIPCKSSPELEDTASS--	2838			
QY	1264	WLRESPPDKA-----	KDYLISIRGLGLKSVCEVRLTLHLNLAFFVDNTVNGRIARVMGWVP	1318			
Db	2839	--	KRRPTRAQKVEVKEELLAVGKLTQTSCE-----	TTHTDKEPVGEGKTKAFK	2886		
QY	1319	LQPLPESLQHLHLELYPVLESIOKFLWPLRCKLDQRTLYEL-----	HYQLITFG	1367			
Db	2887	QPAKRN-----	VDAEDVIGSRQ--	PRAPKEKAQPLEDLASFOELSQTSGHTEELANG	2937		
QY	1368	KVCTSRPNCNACPMRGCRHFASAYASARLALPAPEER--	SLTSATIPVPPPEPPV	1424			
Db	2938	AADSFTSAP--	KOTPSGK-----	FLKISRRLRAPKVEPVGVDVSTRDPVKSQSKSNT	2989		
QY	1425	AIDMIELPLPLEK--	SLAGAPSNRECEP-----	IEE--	PASPGOECTEITESDTEAD	1476	
Db	2990	SLP-----	PLPFKGGKGDGVSVTGKRLRCMPAPEEIVEELPASKKOR-----	VAPRA	3037		
QY	1477	YNEDPDEIPTIKLNIQFOGWTLEHMERNMELQEGDMSKALVALHPTTTSIPTPK-----	1531				


```

Db 3038 RKSSEPVIMKRSYR-----TSAKRIEPAEELNSNDM-KTNKEEHKLQDSVPENKGISLR 3092
QY 1532 LKNISRLRTEHOVEL-PDSHRLLDGMDKRPDPSPYLLALAIWTPGETA----- 1579
Db 3093 SRQDKTEAQOITEVFLAEIRIENRNEKKPKMKTSPK-MDIQNPDGDKARKPIPRDKYVE 3151
QY 1580 -----NSAOPP-EQKCGGKASGKMCFD-----ETCSECNLSREANSQTVR 1618
Db 3152 NKRCLRSARQNESSQPKVAESGGKSAKVLNQNOKGKEAGNSDSMCLRSRKTSPQAA 3211
QY 1619 GTL 1621
Db 3212 STL 3214

RESULT 2
ID ATRX_HUMAN STANDARD; PRT: 2492 AA.
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Colleaux L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=93179111; PubMed=7874112;
RA Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,
RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA Millasseau P., Khrestchatskiy M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.

```

```

RA Pearce A., Chapman J.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RX E2H2 BINDING.
RX MEDLINE=98167853; PubMed=9499421;
RA Cardoso C., Timsit S., Villard L., Khrestchatskiy M., Fontes M.,
RA Colleaux L.;
RT "Specific interaction between the XNP/ATRX gene product and the SET
RT domain of the human E2H2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).
RN [8]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATRX) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN [9]
RP DISEASE.
RX MEDLINE=20213147; PubMed=10751095;
RA Villard L., Fontes M., Ades L.C., Gecz J.;
RT "Identification of a mutation in the XNP/ATRX gene in a family
RT reported as Smith-Fineman-Myers syndrome.";
RL Am. J. Med. Genet. 91:83-85(2000).
RN [10]
RP VARIANT ATR-X SER-1713.
RX MEDLINE=97196774; PubMed=9043863;
RA Villard L., Lacombe D., Fontes M.;
RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
RT without alpha-thalassemia.";
RL Eur. J. Hum. Genet. 4:316-320(1996).
RN [11]
RP VARIANT JM GLN-2131.
RX MEDLINE=96224392; PubMed=8630485;
RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
RL Nat. Genet. 12:359-360(1996).
RN [12]
RP VARIANTS ATR-X.
RX MEDLINE=97467722; PubMed=9326931;
RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA Levin M.L., Masuno M., Nerl G., Pierpont M.E., Slaney S.F.,
RA Higgs D.R.;
RT "Mutations in transcriptional regulator ATRX establish the functional
RT significance of a PHD-like domain.";
RL Nat. Genet. 17:146-148(1997).
RN [13]
RP VARIANT ATR-X LEU-246.
RX MEDLINE=20123062; PubMed=10660327;
RA Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,
RA Greco D., Cardoso C., Fontes M., Ragusa A.;
RT "New mutations in XNP/ATRX gene: a further contribution to
RT genotype/phenotype relationship in ATR/X syndrome.";
RL Hum. Mutat. 12:214-214(1998).
RN [14]
RP VARIANT SHS LYS-1742.
RX MEDLINE=99347960; PubMed=10417298;
RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
RA Prieto F., Fontes M., Martinez F.;
RT "Mutation of the XNP/ATRX gene in a family with severe mental
RT retardation, spastic paraplegia and skewed pattern of X inactivation:
RT demonstration that the mutation is involved in the inactivation
RT bias.";
RL Am. J. Hum. Genet. 65:558-562(1999).
RN [15]
RP VARIANT CWS THR-2050.
RX MEDLINE=99326061; PubMed=10398237;

```

RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
RT Curtis M.;
RA "Carpenter-Waziri syndrome results from a mutation in XNP.";
RL Am. J. Med. Genet. 85:249-251(1999).
RN [16]
RP VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
RX MEDLINE=99219535; PubMed=10204841;
RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Beloune J.,
RA Loei A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
RT "Evaluation of a mutation screening strategy for sporadic cases of
ATR-X syndrome.";
RL J. Med. Genet. 36:183-186(1999).
RN [17]
RP VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
RX MEDLINE=20451413; PubMed=10995512;
RA Wada T., Kubota T., Fukushima Y., Saitoh S.;
RT "Molecular genetic study of Japanese patients with X-linked
alpha-thalassemia/mental retardation syndrome (ATR-X).";
RL Am. J. Med. Genet. 94:242-248(2000).
CC -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC -!- SUBUNIT: PROBABLY BINDS E2H2. BINDS ANNEXIN V IN A CALCIUM AND
CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
CC similarity).
CC -!- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
CC INTERACTING WITH HP1.
CC -!- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1, 2, 3, 4 (SHOWN HERE) AND 5;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
CC SYNDROME (CWS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH
CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME
CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
CC MICROGENITALISM AND EARLY DEATH.
CC -!- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS
CC RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND
CC BILATERAL CRYPTORCHIDISM. DUE TO THE CLINICAL OVERLAP WITH ATR-X
CC SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SFM, MIGHT
CC BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H
CC INCLUSIONS.
CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED
CC MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE
CC MENTAL RETARDATION WITH SPASTIC PARAPLEGIA, MICROCEPHALY, SHORT
CC STATURE AND CRYPTORCHIDISM.
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U72937; AAB49970.2; -;
CC EMBL; U72938; AAB49971.2; -;
CC EMBL; U72935; AAB40698.1; -;
CC EMBL; U72904; AAB40698.1; JOINED.

Query Match 2.5%; Score 225.5; DB 1; Length 2492;
Best Local Similarity 18.8%; Pred. No. 0.00069;
Matches 254; Conservative 187; Mismatches 510; Indels 401; Gaps 55;

QY 27 LEFLNLTPOOKPSKRKRKMPK-VVVEGK---PKRPRKPAELPKV---EGKPRKPK 79
DB 409 LEEDLNSEFRAMDAVNKEKTKHEKVDAKFPETKARKEKPCALEKKDISKEAKLSRK 468
QY 80 -----RKAATQKVKSTGTSA-KKKNLKESTKTPANVGD-----MSNKSPEVT 123
DB 469 VDSEHMHNQVTEORTNKSTGGHEKSKDRKEEPOYEPANTSEDLMDIVSPVPEDI 528
QY 124 LKSRKALNF--DLENPDARQDQSESEIVQNSGANSFSEIRDAIGCTNGSFIDLSVSI 181
DB 529 FENLETAMEVQSSVDHQDGGSG-TEQEVESVVKLNIS--KDNRG--KSKTTA 580
QY 182 DKTNGLGAMNOPLEVMGNQDPDKLSTGAKLARDQO-----DLITNQCQCPVATQNTQF 237
DB 581 KVTKELYVKLTP--VSLPNSPIKIGADCOEVPQDKGYKSCGLNPKLEKCGLGQENSDNEH 638
QY 238 PMENQOAWLQMKNLGIFPGNOOP-----RMTIRNOQPC-LAMGNQOQPMYLIGTPR 288
DB 639 LVENEVSILLLESDLRSPRVKTTPLRPTETNPVTSNDEECNETYKEKOKLSVPVRKK 698
QY 289 PALVSGNOQLGGPOGNRPP-----IF 309
DB 699 DKRNSDSASIDNPKNPKLPSKQSETVDQNSDSDEMLAILKGVSRMSSSSDITNEIH 758
QY 310 LNHOTCLPAGNQLYGSPTDMHQLVMSTGGQOGLLLRNQOPGSLIRGOQPCVPLIDQO 369
DB 759 TNHKTLDVLTQAGKDDKGRKRKSSSTSGSDFDTKKGKSAKSSIISKK-----RQTS 812
QY 370 TPKGFTHLNQMVATSMSSPGLRPHSQSOVPTTYLHVESVSRILNGTTGTCORSRAPYDS 429
DB 813 ESSNYDSELEIKSMKIGAAATTKRIPNT-----KDFDS 849
QY 430 LQODIH-----QGNKYI-LSHEISNGCKKALPQNS-----LPTPIM-----A 468
DB 850 SEDEKHKMGMDNGHKNLKTQSEGSDDAERKQERETFSAGTVDKDTTMELRDLRP 909
QY 469 KLEARGSKRQYHRAMGQTEKHDLNLAQIAQSDQVDEHNSSTCVLEYLDAAKYTKIQKV 528
DB 910 KQOASASTDGVKLSKEQSFSTLEVRKVAETKEKSKH-----LKTCTCKV 957
QY 529 QENLHGMPPEVI-----EIEDPTDGARKG-----KNTASIKGASK-----GNSSPVK 572
DB 958 QDGLSDIAEKFLKDKQSDSETSEDKKQKGTETBKPKPSDFKKVKMEQOYESSSDGTE 1017
QY 573 KTAKEK-CIVPK--TPAKKGRA-GRKSVPPPAHASEIQLWQTPPKTPLS-----RSKP 624
DB 1018 KLPEREETCHFPKGIQTKNGTTDGEKSKIRDKTSK-----KKDELSYAEKSTG 1069
QY 625 KGGRKSTODSGKARGSPGELLQDSIAEIIYRMQNLVLDGKEREQONAMVLYKGDGAL 684
DB 1070 KGDSCDSEDKSKNGAYGR-----EKKRCK-----L 1096
QY 685 VPYSEKRRKPRPKVIDDETTRINWLLMGKDEGEDEKDKKKEKWEERVRFRGRAD 744
DB 1097 LGKSSRRHQ-----DCSSSDTEKYSM-----KEDGCSNDRKRLRIERRNLSKRN 1145
QY 745 SFIARMLVQGDREFSPWKGVSVDVIGVFTQNVSDHLSSAFMSLAARPPKLSRSRE 804
DB 1146 -----TKEIQSGSSSS-----DAEESSE 1163
QY 805 DER-----NVRSVVVEDPEGCLILNLEIPSMOEKVOHPSDMEVSGVDSKSEQLRDC 856
DB 1164 DNKKKQRTSSKKKAVIVKEKK-----RNSURTSTKRRQADITSSSSDIEDD 1211
QY 857 SNSGIERNFLEKSTQNLEEE-VLSSQDSFDPAPFQSGRVRGSCSCSKS-----DA 906
DB 1212 DQNSIGEGSSDEQKLPVTENVLSSHTGF-----CQSSGDEALSKSVPTVDDDDDD 1264
QY 907 EFPTRCTKTVSGTSQSVQTSQSPNLSDICLOGNERPHLYEGSGDVQ----KQETTIVA 962

```

Db 1265 NDENRIAKWML-----LBEIKANLSSD-----EDGSSDDEPEGKRTQKQ 1307
QY 963 QKRPDLKTNWKLDSVCFQGRNDTNMQTTPSSSYEQCATRQPHVLIDIEFGMGQGLGY 1022
Db 1308 EENPGDEAKNQVNS-----ESDSSEESKKPYRHRLLR--HKLTVSD-GESE----- 1354
QY 1023 SWMSISPRVDRVANKNVPRFFRGGSVPREFTGQIIPSPHPLPGMGLSGSSSAVOEHQ 1082
Db 1355 -----EKKTKPKHKHVKVGNRRK-----VSSDESDDSDQFQSGVSEEVSESE 1397
QY 1083 DDTQHNOD-----ENMKASHLOKTFDLNLSSECLTRQSSTKQNTDGCCLPRDRTAE 1136
Db 1398 DEORPRTSRKAKAELEBNQSYQK-----KKRR-----RIKQVE 1432
QY 1137 DVVDPLSNSSLONLIVSNESNKKQTAVEYKETAATILREMGKTLADGKKPTQOWDLSR 1196
Db 1433 D-----SSSEN---KSNSEEEEEKSEEEEEEEEEDEEDDSDSKSPGKRKKIR 1481
QY 1197 KVEGNEGROERNKNNMDSIDYRAIRASISE 1228
Db 1482 KILKDDKLRT-----TONALKKEEERKKRIAE 1509

RESULT 3
MAPB_HUMAN STANDARD; PRT; 2468 AA.
AC P46821;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
DE LC1].
GN MAP1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95104835; PubMed=7806212;
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RT Cloning of human microtubule-associated protein 1B and the
RT identification of a related gene on chromosome 15.;
RL Genomics 22:273-280(1994).
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
CC -1- SIMILARITY: TO MAP1A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL: L06237; AAA18904.1; -.
CC MIM: 157129; -.

```

```

DR InterPro: IPR000102; MAP1B_neuraxin..
DR Pfam: PF00414; MAP1B_neuraxin; 10.
DR PROSITE: PS00230; MAP1B_NEURAXIN; 6.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
FT REPEAT 1878 1894 MAP1B 1.
FT REPEAT 1895 1911 MAP1B 2.
FT REPEAT 1912 1928 MAP1B 3.
FT REPEAT 1929 1945 MAP1B 4.
FT REPEAT 1946 1962 MAP1B 5.
FT REPEAT 1963 1979 MAP1B 6.
FT REPEAT 1997 2013 MAP1B 7.
FT REPEAT 2014 2030 MAP1B 8.
FT REPEAT 2031 2047 MAP1B 9.
FT REPEAT 2048 2064 MAP1B 10.
FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 2.5%; Score 225; DB 1; Length 2468;
Best Local Similarity 18.5%; Pred. No. 0.00072;
Matches 336; Conservative 254; Mismatches 710; Indels 518; Gaps 81;

QY 175 LDSVSQIDKT-----NLGAMNQPLEVSMGNQDPDKLSTGAKLARQQDQLLTRNQCCQFPV 230
Db 301 LDRVDSILLTHIGDDNLPGINSMLQRIAELEEBEQSGSTNSDMKNLISDPLGVVFLN 360
QY 231 ATQNTQFPMENQQAWLQMKNLQIGFPGFNQOPRMTIRNQOPCL-AMGN----- 277
Db 361 VPENLKNPEPN-----IKMKRSIEACFTLOYLNKLSMKPEPLFRSVGNTIDPVILFOKMG 416
QY 278 --QOPMYLIGTPPA--LVSGNQQLGG-----PQGNKRPIFLNHTCL----- 316
Db 417 VGKLEMLVLPNVKSSKEMOYFMQWGTCTNKDKAEFILPNGQEVDPISYLTSSVSLVNH 476
QY 317 PAGNQLVGSPTDMHQLVMSTGGQ-OHGLL-----IKN-----QOPGSL---IRGQOPCVPL 363
Db 477 PA-----NPAEKIIRVFPNGNSTQYNILEGLEKHLKHLDFLQPLATQKDLTGQVP-TPV 529
QY 364 IDQOPATPKGFTHLNOM-----VATSMSSPGLPHSOSQVP--TTLVHVESVSRL 412
Db 530 VKQ-----TLKQRADRSRESLKPAKPLPSKVSRRKSEKTEPPTVKVNHVE----- 575
QY 413 NGTGTGQSRAPAYDSIQDIIHQGNKYILSHSISNGCGCKKALPQNSSLPTPTIMAKLEE 472
Db 576 -----KPKVESKEKVMVKDKDPKTE--TKPSYTEKEVPSKEE-PSPVKAEVAE 622
QY 473 ARGSKROYHRAMGQTEKHDLNLAAQIIAQSO-----DVERHNSSTCVVEYLDAAKTKIQK 526
Db 623 KOATDVKPKAAKEKTKVKRPEDKKEKEKPKKPAKVEDKTPIKKKEKPKKEEVKK 682
QY 527 VQENLHGMPEVIEIEDDPTDGARKCKNTASISKGASKGNSSPVKKTAEKCIIVKTP 586
Db 683 EVKKEIKK-----EKEKPKKEVKKETPPKPKVEKKEKKEKKEK-EEKE-----PKKE 731
QY 587 AKKGRACRKKSVPPPAHASEITQLWQTPPP-----KTPLSRSKPKGKGR-KSIQDSGKA 638
Db 732 IKKLPKDAKSSSTPLSAPKPAALKPKVPKKEESVKKDSVAAGKPKGKGIKVIKKGKA 791
QY 639 RGPSEGLLCODSTAEIITYRMQNYLGDKEREQONAMVLYKGDGALVPEESKKRKRPKV 698
Db 792 AEAVAAAVGTGATTAAVMAAAAGIAAIGPAKELEAERSLSMSSPEDLTQDFELKAE---EV 848
QY 699 DIDDETRINWLLMGKDEGEDEKDKKEK-----WWEERRVFRGRADSFIAHMLVQ 754
Db 849 DVTKDIPKQLELI-----EDEEKLKETEPVEAYVIQKEREYTKGPAESDGEITTE 900
QY 755 GD-----RRFSPWKGVSDVSVIGVELTQNVSDHLSSSAFMSLAARPPKLSRSSREDER 807
Db 901 GEGECEQTPPELEPVEKQGVDDI-----EKFEDEGAGF-BESSETGTDEE 944
QY 808 NVRSVVVVEDPEG-----CILNLNEIFSWQEK-----VQHPSDMEVSGVDSGSKEQ 852

```


or send an email to license@isb-sib.ch).

CC EMBL; AB029290; BAA83821.1; -
CC EMBL; AF141968; AAF06360.1; -
DR EMBL; AB033077; BAA86565.1; -
DR EMBL; AL137853; CAC15920.1; -
DR EMBL; AB007934; BAA32310.1; -
DR HSP; Q01082; IBRR.
DR InterPro; IPR001589; Actinin_act_bind.
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR003108; GAS2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00306; ehand; 2.
DR Pfam; PF02187; GAS2; 1.
DR Pfam; PF00435; spectrin; 35.
DR SMART; SM00033; CH; 2.
DR SMART; SM00054; Efh; 2.
DR SMART; SM00243; GAS2; 1.
DR SMART; SM00150; SPEC; 35.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS50021; CH; 2.
DR PROSITE; PS50018; EF_HAND; 2.
DR PROSITE; PS50002; SH3; FALSE_NEG.
KW Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain.
FT Actin-binding; 1 295
FT DOMAIN 1 181
FT DOMAIN 194 295
FT REPEAT 314 355
FT REPEAT 591 623
FT REPEAT 680 784
FT REPEAT 786 800
FT DOMAIN 871 923
FT REPEAT 1250 1272
FT REPEAT 1287 1342
FT REPEAT 1455 1534
FT REPEAT 1547 1659
FT REPEAT 1815 1891
FT REPEAT 1932 2042
FT REPEAT 2260 2280
FT REPEAT 2372 2395
FT REPEAT 2398 2507
FT REPEAT 2510 2618
FT REPEAT 2621 2728
FT REPEAT 2731 2838
FT REPEAT 2841 2945
FT REPEAT 2987 3024
FT REPEAT 3136 3163
FT REPEAT 3187 3274
FT REPEAT 3277 3383
FT REPEAT 3386 3492
FT REPEAT 3495 3601
FT REPEAT 3604 3673
FT REPEAT 3713 3819
FT REPEAT 3832 3927
FT REPEAT 3982 4043
FT REPEAT 4046 4152
FT REPEAT 4155 4262
FT REPEAT 4265 4371
FT REPEAT 4374 4481
FT REPEAT 4484 4590
FT REPEAT 4593 4700
FT REPEAT 4707 4808
FT REPEAT 4811 4917
FT REPEAT 4920 4985
FT REPEAT 5030 5054
FT CA_BIND 5096 5107
FT CA_BIND 5132 5143
FT CA_BIND 5276 5283
FT DOMAIN 5355

4 X 4 AA TANDEM REPEATS OF [GS]-S-R-[AR].

FT CONFLICT 1 72 MSSDETLSSRSRSCRSRSYRSRSGSLSPCPGPD
FT FT TLPWNLPHEQKKRKSQVLDPAERAVRV -> MFDVLW
FT FT AGIPGRDVGSLQPLPPGFKQKTSASRAVI (IN REF.
FT FT 2).
FT FT VAISSEDEGNLRFVVELLS -> GPSAPLKMKAISDLCLMN
FT FT YCL (IN REF. 1).
FT FT A -> T (IN REF. 2 AND 3).
FT FT V -> A (IN REF. 1).
FT FT E -> D (IN REF. 1).
FT FT E -> K (IN REF. 2).
FT FT M -> V (IN REF. 2).
FT FT C -> Y (IN REF. 1).
FT FT MISSING (IN REF. 2).
FT FT SILPSVG -> EYRLFKI (IN REF. 3).
FT FT Q -> R (IN REF. 2 AND 4).
FT FT S -> T (IN REF. 2).
FT FT MISSING (IN REF. 2).
SQ SEQUENCE 5430 AA; 620346 MW; 91ADB7F7580B440B CRC64;

Query Match 2.4%; Score 219.5; DB 1; Length 5430;
Best Local Similarity 17.5%; Pred. No. 0.0041;
Matches 323; Conservative 296; Mismatches 702; Indels 527; Gaps 78;

QY 16 TEQNDGSRDVLFEFLDKTPQOKSKRKRKMPKV-----VVEGPKRKRKPAELPKV 69
Db 1577 TTQDLSALQKNQSDL-KDLQDDIQNRATSFATVVKDIEGFMEENOTKLSPRELTAL--- 1632
QY 70 VVEGPKRKRKAATQEKVKSKETGSAKKKLNKESATKKPANVGDMNSKSPVTLKSCR 129
Db 1633 -----REKLHOAKEOYEAQOEETRAQKE-LEEAVT-----S 1663
QY 130 ALNFDLENPGDARQGDSESEIVQNSSGANSFSEIRDAIGTNGSFLDSVSOIDKTNGLGA 189
Db 1664 ALQOETEKSKRAK-----ELAENKKIDALDWTSVGSSGQLLTNLPGMEQLSGASL 1717
QY 190 MNQPLEVSMG-----NQ-PDKLSTGAKLADQOPDLLTRNQOCQFPVATONTOPFMEQQA 244
Db 1718 EKGALDITDGVMGVNAPEKLDKQCEMMKARHOELL--SQQNFILATQSAQAFLDQGH 1775
QY 245 WL-----OMKNOLIGFPFGNQOPRMTIRNQOCLANGNQOPMYLIGTPPALVS----- 293
Db 1776 NLTPEEQMLQQRKG-----ELKEQYSTSLAQSEAEKQVOTLODELQKFLQDHK 1825
QY 294 -----GNQQLGGPQGNKRPIFLNHQTCPLPAGNQLYSGPTDMHQLVMSTGGQQ 340
Db 1826 EFESWLERSEKLENNHKGSSPETLPSLLKQ-----GS---FSEDVSHKGD 1872
QY 341 HGLLKNOQ-----PGSLINGQPCV---PLIDQOPATPKGFT-----HLNQMVA 382
Db 1873 REVTTISGQKVLDMENSFKEGKEPSEIGNLVKDKLDATERYTALHSKCTRGLSHLMLIG 1932
QY 383 TSMSSPGLRPHSQSOVPTYLHVESVSRILNGTTCQORSRAPAYDSLOODIHOQNKYL 442
Db 1933 -----QYHQFQNSADSLQAMQACEAN-----VEKLSDTVADSDPGVL 1970
QY 443 SHEISNGNGCKKALPQNSLPTPIAK-----LEEARGKROYHRAMGOTEKHDLNLAQOI 498
Db 1971 QBOLATTQLOEELAEH-QVPVEKLOKVARDIMELEGEPADPHRHVQETDLSILSHFOSL 2029
QY 499 AQSDQVERHNSSTVEYLDAAKTKIQV-----VQENLHGMPPVEVIEIDDPDGDGARK 552
Db 2030 S-----YSLAERSLLQRAIAQSQSVQESLESLSLQSIGVEQN-----LE 2069
QY 553 GKNATASIKGASKNGSSPVKKTAEKICIVPKTPAKKGRAGRKKSVPFPAHASEIQLNQ 612
Db 2070 GKQVSSLSGVI-----OEALATNMKLDIARQKSSL--EATREMYTRFME 2114
QY 613 TPKPTPLSRSPKGGKRSIQDSGRARGPSGELLQODSTAEIYRMONLYGDKEREQEQ 672
Db 2115 TADSTAA-----VLOKLAEVESQRFEQLCLOQK----- 2145
QY 673 NAMVLYKGDGALVPYSEKRRKRPKVDIDDETT-RIWNLMLMGKDEKDEKKEKW 731

Db	2146	-----ESSLKKLLFOAEFMFHLGSKLQOQFNMENKSRMLASQNPQDDITHF	2190
Qy	732	WEERRVYRGRADSFIAHMLVQGDREFPWKGVSVDSVGVFTQNVSDHLSASFMSL	791
Db	2191	FOQ-----IOELNLEMDQOEN-----LDTL-----EHLVTELSGCFALD	2226
Qy	792	AARFPKLSRREDERNVRSVVE---DPEGCILNLE-----IPSWQEKVQ---HPS	838
Db	2227	LCQHQRVONLRKDFTELQTKVREKENDASCCQQLDFFERKLVRTFOKWLKETEISPT	2286
Qy	839	DMEVSGVD-----SGSKEQLRDCNSG---TERFNFLEKSNLNEEVLSSQD-----SFD	886
Db	2287	ETMSAKELEKQIEHLKSLDDWASKTLVEEINCKGTSLENLIMEITAPDSQKGTGSIL	2346
Qy	887	PAIFQSGCRVG---SC-----SCSKSDAEFP-----TTRCET---KT	917
Db	2347	PSVGSVGSVNGYHTCKDLTIQICDMSDVNLKYEKLGVLHERGESLQAILNRMEEVHKE	2406
Qy	918	VSQTSQSVOT-----GSPNLDEICLOGNER-----PHLYEGSGDVQ---KQETT	960
Db	2407	ANSVLQWLESKEEVLKSDAMSSPTKETVKAQAESKAFLAELQNSPKIQVKKEALAG	2466
Qy	961	VAQKKPDLKTMNWKDSVCFQPPRNDNW-----QTPSSSYEOC-----ATR	1003
Db	2467	LLVTYPNSQEAENKKI-----QEELNSRWRATEVTVARQQLSEASHLACFOAAESQL	2522
Qy	1004	QPHVLDIE-DFCMOGEGLYSWMISPRVDKVNKNVPRREFRGGSPREFTGQIIPST	1062
Db	2523	QPLWKEKLMGVILGP-----LSDPMLNAQKQV---QWMLKEFFARRQHQEQLNEA	2574
Qy	1063	PHLEPGMG-LSGSSSAVOEHODDTQHNOODEMNKASHLQKTFLLDNLNSEECLTRQSGTK	1121
Db	2575	QGILTPGQVSLSTSVOKE-----LQSIQKWVELTDKLSNR	2612
Qy	1122	QNITDGLPDRORTADVDDPLSNSSL---QNILVESNNKEQTAVEYKETAAILREMK	1179
Db	2613	SSQIDQATVKSTQYQELLQDLSEKVRVAGORLSVQSAISTQPEAVKQOLETSEIRSDLE	2672
Qy	1180	GTADGKPTQWDSLRKDVNEGROBRKNMDSIDYAIRRASISEISEAIKERGMN	1239
Db	2673	QLDHEVKAQFLCDBLSVLI-GEQYKDELKRLKLETV-----ALPQGLE	2716
Qy	1240	NMLAVRIKDFLERIVKDHGGIDL-----EMLRESPDKAKDYLLSIRGLGKS-----VE	1289
Db	2717	DLAADRLNQLAALASTQOFOQMFDELRTLWDDKQSQQAKNCPISAKLERLSQLOQNEE	2776
Qy	1290	CVRLTLHLNLAFFVDTNVRTAVRMGWYPLQPLPESLQLHLELYPVLESIOKFLWPRLC	1349
Db	2777	FOKSLNQHSYGSEVIVAEGESLLLS--VPPGEKRTLQNLQVELKNHWHEELSKKTADRQS	2834
Qy	1350	KL-----DORTLYELHYQLITFGKVCTKSRPNCNACPMRGCRHPASAYASARLALPAPE	1405
Db	2835	RLKDCMQAKQYQWHE-----DLVPWIEDCK-----AKMSRLRVLTD---2872	
Qy	1406	ERSLTSATIPVPPEFPVPAIMIELPLPLEKS-LASGAPSN-----RNCPEPIEERASP	1460
Db	2873	-----PVQLESLLRSKAMLNNEVEKRRSLLEILNSAADI	2906
Qy	1461	GOECTETTESDIED--AYNEDPD---EIPIKLNIEQFGMTLEHME--RNMELQ-EG	1511
Db	2907	LINSSEADEGHRDEKAGINONMDAVTELOAKTQGSLEEMTQRLREFQESFNIEKKVEG	2966
Qy	1512	DMSKALV--ALPHTTTSIPTPKLNISRLRTEHQVVEYELPDS-----HRLLDGMDKREP-	1563
Db	2967	AKHQLEIFDALQSQACS-----NKNLEKRAQOEVLQALEPOQVDYLRNFTQGLVEDAPDG	3021
Qy	1564	-DPSYLLAIWTPGETANSAPQPEQKCGKASGKWCDFETCSECSNLSREANSQTVRGTL	1622
Db	3022	SDASQLL-----HQAEOVAQOE-----FLEVQKRVNS-----GCVN	3051
Qy	1623	IFCRTAMRGSFFLNGTYFQVNFLEFADHESLKPIDVPRDWIDLPRT	1670

Db	3052	MENKLEGIGOF-----HCRVREMF-----SQLADLDDLDGCGMGAIGRDT	3090
RESULT	5		
APC_RAT			
ID	APC_RAT	STANDARD;	PRT; 2842 AA.
AC	P70478;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Adenomatous polyposis coli protein (APC protein).		
GN	APC.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=101116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FISCHER 344/N; TISSUE=Brain;		
RX	MEDLINE=961116966; PubMed=8563176;		
RA	Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,		
RA	Sugimura T., Nagao M.;		
RT	*cDNA cloning of the rat APC gene and assignment to chromosome 18.;		
RL	Mamm. Genome 6:746-748(1995).		
RN	[2]		
RP	MUTAGENESIS.		
RC	STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;		
RX	MEDLINE=95148647; PubMed=7846077;		
RA	Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,		
RA	Weisburger J.H., Sugimura T., Nagao M.;		
RT	Tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.;		
RT	Specific 5'-GGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon		
RL	Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).		
CC	-1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-		
CC	CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE		
CC	ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY		
CC	SIMILARITY).		
CC	-1- SUBUNIT: FORMS HOMOLOGOMERS AND ASSOCIATES WITH CATENINS (BY		
CC	SIMILARITY).		
CC	-1- PTM: PHOSPHORYLATED BY GSK-3b (BY SIMILARITY).		
CC	-1- SIMILARITY: CONTAINS 7 ARM REPEATS.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; D38629; BAA07609.1; -.		
DR	HSSP; Q02248; 3BCT.		
DR	InterPro; IPR000225; Armadillo.		
DR	PFam; PF00514; Armadillo_seg; 6.		
DR	SMART; SM00185; ARM; 5.		
DR	PROSITE; PS50176; ARM_REPEAT; 1.		
KW	Anti-oncogene; Phosphorylation; Coiled coil; Repeat.		
FT	DOMAIN 1 728		
FT	DOMAIN 1 62		
FT	DOMAIN 125 260		
FT	REPEAT 451 493		
FT	REPEAT 503 545		
FT	REPEAT 546 589		
FT	REPEAT 590 636		
FT	REPEAT 637 681		
FT	REPEAT 682 723		
FT	REPEAT 724 765		
FT	DOMAIN 739 2831		
FT	DOMAIN 1130 1155		
FT	DOMAIN 1556 1575		
FT	DOMAIN 1864 1891		
FT	MUTAGEN 523 523		
FT	SEQUENCE 2842 AA; 310530 MW; 3CB2EA8A34E8F47 CRG64;		

Query Match		2.3%; Score 211.5; DB 1; Length 2842;
Best Local Similarity		18.4%; Pred. No. 0.0044;
Matches		319; Conservative 230; Mismatches 654; Indels 533; Gaps 83;
Qy	64	AELPKVVVE-----GPKPKRPAKAAOEVKVKSETSAKKKLIKESATKKPANV-----112
Db	883	AQIAKVMEEVSALHTSQDORSPASAAELHCAVEERTAA-----RSSASHTPHNFNAK 937
Qy	113	GDMSKKPEVTL-----KSCRKALNFDLENPGDARGDSESEIVQNSGANSSEIRDA 166
Db	938	SESNKRCMPYAKVEYKRSNDSLNSVTSDGYGKRG-----QMKPSVESYSEDEG 990
Qy	167	IGGTNGSF-LDSVQIDKTGL-----GAMNOPLEVSMGNOPDKLSTG-----AKLARD 214
Db	991	KFCVSGVQYPADLAHKIHSANHMDNGGELDTPINSLKYSDEQLNSGRQSPQNERWARP 1050
Qy	215	QO--PDLLTRNQOQFPVATONTQFPHNQOAWLQMKNLQIGLFFGNOQPRMTTRNQOPC 272
Db	1051	KHVIEDEIKONEQRQ--SRSQNTNFPVYSENT--DDKHLAKFOOHFG-----QOEC 1096
Qy	273	LAMGNQOPMYLIGTPRPALVSGNQLGPGQGNKRPILNHOVC-----LPAGNQLYG 324
Db	1097	VS-----PYRSRGTHG-----SETRMGSSHAVNQNV--NOSLCOEDDYEDDKPTNYSEY 1145
Qy	325	SPTDMHQ-----LVMSTGGQHQGLLIKKNQP-----GSLIRGQQPCVPLIDQQA 369
Db	1146	SEEQHEEERPTNYSIKYNEEKHV-----DQPIDYSILKYATDIDSSQOKPSFS-FSKTSP 1200
Qy	370	TPKFTHLNQWVATSMSPGLRHQSQVQPTTYLHVSVSRILNGTGTQORSRAPAYDS 429
Db	1201	VQGTKEHNSPSSAASAPSSNARQSQ-----LHPSSAQR--NGQTPKGTACKVP--\$ 1250
Qy	430	LOQDIHGNKYILS-----HEISNGN--CCKKALPONSLPTPIMAKLEEAR 474
Db	1251	INQETMO--TYCVEDTPICFSRCSLSLSAEDIEICDQDTQCADSANTLIQIAEIKEND 1308
Qy	475	GSKROYH-----RAMGQTEKHDLNLAQIAQSDQVERHNSSTVCVEYILDAK--KTKIQK 526
Db	1309	VTRSAQDPASDVPAVQSTRTKPSRLQASGLASESARHKA--VEFSSGAKSPSKSAQT 1365
Qy	527	VQENLHGMPEVIEEDDPTDGARKGNKTASISKGASKNSPVKKTAEKCI-----581
Db	1366	PKS-----PPEHY-VQETPLVFSR-----CTSVS-SLDSFESRSIASSVQSEPCSGMVSG 1413
Qy	582	-----VPKTPAKKGRAGKRSVPPRAHASIQLWQPTPKTPLSRKPKGKGRKSTQDS 635
Db	1414	IVSPSDLPDPSGQTMPPRSKTPPPP-----PPQPVQTKREVPKTKVPAABQRE 1463
Qy	636	GKARGSGELLCQDSIAEIIYRMQNLVLDGK-----EREQEQNAMVLYKGDGAL---VPYE 688
Db	1464	G--GPK-----QTAVSAQVQVLPDADTLHFATFESTPDGFCSSSLSALSLEPFI 1515
Qy	689	SKKRRPR--PKVIDDDETRINLLMGKGB-----KEGDEKDKKKKEKWEE-----735
Db	1516	QKQVELRIMPPVQEND-----NGNETEPEQESNENQDKVEKPDSEKDLDD 1564
Qy	736	-----RRVFR-----GRADFSFIARMLHVQDORR---758
Db	1565	SDDDIEILEECIISAMPTKSSRAKAKLAQTASKLPPPVARKSQLPVYKLLPSQSRLQA 1624
Qy	759	-----FSPWKSGVVDSVIGVFLTNQVSDHLSAFAFMS-LAARPPPKLSSSRDEDERNRSV 812
Db	1625	QKHVSFTFG-----DDVPRVYCVEGTINSTATSLDILIESPNELAGD---GVRAS 1676
Qy	813	VVEDPEGICILNLNIPSQEOKVQHPDSMEYSGVD-----SGSKEQ-----LRDCNSGIERF 864
Db	1677	V--QSGEFEKRDPTIPEGRSTDEAQRKVSSIAIPDLDGSKAEGLDILAEICNSA----1729
Qy	865	NFLKSTQNLBEEVLSQDSDPDPAIFOSG-----RV 896
Db	1730	--LPKGRSHKPFVKIMDVOQOASMTSSGNTKNQIDTKKKKPTSPVKPMPQNTYRTRV 1787
Qy	897	GSCSKSDAEFPTRCTKTIVSGTSQSVQTSQVGNLSDEICLOGNE-----RPHL 946
Db	1788	RKNTDSKVNVNTEETFSDNK--DSKQSLKNNPKDLNDK--LPDNEDVRGGFTFDSPHH 1843
Qy	947	Y-----EG-----SGDVOKQETTNVAQKKPDLEKTMNMKDSVCFQOPRNDTNQOT 991
Db	1844	YATIEGTPYCFSRNDSLSLDFD-DDDVLRSREKAELRKGEKSDS-----EAKVTCHT 1896
Qy	992	TPSSSYE-----OCATROPHVLDIEDFGMOGELGYKSWMSISPRDVRVKNKVPRRFRQ 1046
Db	1897	EPSSQOSAKKAQASTKHP-----VNRGPKPLLOEQPTFPO-SSKQVDR-----1941
Qy	1047	GGSVPREFTGQIIPSTPHELPGMGLSGSSSAVQEHQDDTOHNOQDENMKASHLQKTFDL 1106
Db	1942	GAATDEKLQFAIENTP-----VCFSRNSSLSDVDQENNNNEETGPV-----1986
Qy	1107	LNSEECLTQSSSTKQNTGCGCLPRDRTAEDVDVPLSNSSSLQNLILVESNSSNKEQTAVE 1166
Db	1987	--RDAEPANAQGGQPGKPAQSGYAPKSFHVEDTPVCFSRNSSLSDSDDED-----2036
Qy	1167	YKETAATILRE-MKGTLDGKKPT-----SOWDSL-----KQVEGNEGRO 1206
Db	2037	-----LLRECISAMPKRRPSRLKGEQWQSPKVGSVLAEDLTLDLKDQRPSEH 2089
Qy	1207	ERNKNNMDSIDYEAIRRASISEIAIKERGMNMLAVRIKDFLERIVKDHGIDLEWL 1266
Db	2090	GLSPDS-ENFDKWAIEGANSIVSSLHQAAAAACLSQAASSSDSILSLKSGVSLGSPF 2148
Qy	1267	ESPPDKAKDYLSTIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGVPLQPLPESL 1326
Db	2149	HLTPDQEEKPTSHKG-----2164
Qy	1327	QLHLELYPVLESIQKFLWPLRCLKDORTLYELH-----YOLITFGKVFC 1371
Db	2165	-----PRILKPGKSTLEAKKIESENKKGKGGKVKSLITGKI--2203
Qy	1372	TKSRPNCN-ACPMRGECRHFASAYASARLALPAPEERSLTSATIPVPPSPVAIPMIE 1430
Db	2204	--RSNEISSQMKQPLQTNMPSISRGRTMIHIPVGRNSSSTSPVSKKG-PP-----2252
Qy	1431	LPPLPLEKSLAG--APSNRENCEPIIEEPASQOECTEITESDIEDAYNEDP-----D 1482
Db	2253	LKTPASKSPSEGPVATTSPRGTKPAVKSELSP-----ITROTSHISGSKNGPSRSGRD 2306
Qy	1483	EIPTIKLNIQFGMTLREHMERNMELQEGDMSKALVALHPTTTSIPTP-KLKNISR 1537
Db	2307	STPS-----RPTQOPLSRPMQ-----SPGRNSISPGRNGISTPNKLSQLPR 2347
RESULT 6		
NEST_RAT		
ID	NEST_RAT	STANDARD; PRT; 1805 AA.
AC	P21263;	
DT	01-MAY-1991 (Rel. 18, Created)	
DT	01-MAY-1991 (Rel. 18, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Nestin.	
GN	NES.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=90150286; PubMed=1689217;	
RA	Lendahl U., Zimmerman L.B., McKay R.D.G.;	
RT	"CNS stem cells express a new class of intermediate filament	
RT	protein";	
RL	Cell 60:585-595(1990).	
CC	-/- TISSUE SPECIFICITY: CNS STEM CELLS.	
CC	-/- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN	
CC	IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.	

CC CC -|- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; M34384; AAA41685.1; -;
DR PIR; A34736; A34736.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 2.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone.
FT DOMAIN 1 HEAD.
FT DOMAIN 8 314 ROD.
FT DOMAIN 315 1805 TAIL.
FT DOMAIN 8 43 COIL 1A.
FT DOMAIN 44 55 LINKER 1.
FT DOMAIN 56 151 COIL 1B.
FT DOMAIN 152 174 LINKER 12.
FT DOMAIN 175 193 COIL 2A.
FT DOMAIN 194 196 LINKER 2.
FT DOMAIN 197 314 COIL 2B.
SQ SEQUENCE 1805 AA; 198744 MW; B40EE14717E0998D CRC64;

Query Match 2.3%; Score 210.5; DB 1; Length 1805;
Best Local Similarity 18.7%; Pred. No. 0.0027;
Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

QY 60 PRKPAELPKVVEGPKRKRKAATQEKVSK--ETGSAKKNNLKEKATKKPANVGDMN 117
DB 159 PRRP-----PAPHRIPGAPVEEDLARLGEVWRGAVRDYQERYVAHVHESLSGQARE 210
QY 118 KSPVET--LKSCRKALNFLENPGDARGDSSEIVONSSGANSFSEIRDAI--GGTNGSF 174
DB 211 RLSQAVRGARECR-----LEVVQLOADROSLQRERALEORLSGRW 251
QY 175 LDSVSDQDKTN--GLGANNQPLEVSMGNQPKLSTGAKLARDQDPDLTRNQOCPFPVATQ 233
DB 252 QDRLOATDKFOLAVEAEQEKQGLSQIAQLGEGQGLAH-----LKMSLSEVATY 303
QY 234 NTQPFMENQAWLQMKNLGIFPGNQPRMTIRNQOPCLAMGNQPMYLGTPRPALVS 293
DB 304 RTLEAENSR--LQTPGRGSAISGLFDPKL-----RPNF--LGIPEDQYL--GSVLPALSP 354
QY 294 GNQQLGGPQGNKRPI--FLNHQTCLPAGNOLYGSPTDMHOLVMSTGGQHGHLIKNQOPG 351
DB 355 TSFSPPLNTLETPTVTAFLKTQEFLOARTPTLAS--TPIPPISEAP-----CPN 402
QY 352 SLIRGQPCVPLIDQOATP---KGFTHLNMVATSMSSPG-----LR 391
DB 403 AEVRAQVPLSLQTOAPEPLWKATVPSSAILPELEPFGKQKQGHFPDDLTSLATNLN 462
QY 392 PHSQSQVPTTTLHVESYSRIILGTGTCQSRAPAYDSLOQDIHGNKYI-----LSHEI 446
DB 463 PHH-----PT-----LEAKDGSSRSRVSSI--FQEDGQIWEIWEKEADIEVKV 505
QY 447 SNGNGCKKALPQNSSLPTPIMAKLEAGS-----KROYHRAMGQTEKHDLNLAQOIAQSQ 502
DB 506 ENSSAQK---TQESGLDT-----EETQDSQGPLQKETLKAIG--EPLMSLKIQNYETA 554
QY 503 DVERHNST-----CVEYLDAAKTKTKQVVQENL-----HGMP--PEVIEDDPTD 548
DB 555 GKNCNSSTEGHGLTLEGPEKEKQIPILKSLFEKNVSEKTLENGVPVLSLLGKEDRTTE 614
QY 549 -----GARK-----GKNTASTSKGAKGN-----SSPVKKT 574
DB 615 DOELMSPGTLKRFSSLGKSEQEVVVRFSKGNLBSWTAFKEESQHPGLGFGCAEDQMLERL 674

QY 575 AEKEKIVPKTPAKKG-----RAGRKKSVPPPAH-----ASEIQLWQPTPPK 616
DB 675 VEREDQSFRSPSPEEDQACRPLQKQNEQPLGYEEAEGQILERLIEKESQESLRSPEED 734
QY 617 TPLSRAPKPKG----GRKSTQDSGKARGPSGELLQCQDSIAEIIYRMQNLVLDGK-EREQE 671
DB 735 QEAGRSQKENQEPPLGYEEAEDQMLER-----LIEKESQESLSPENRIGKPLERENQ 789
QY 672 ONAMVLYKGCALVPYSEKSKRPKPRKVDIDDETTRIWNLLMGKDEKGD--EEKDKKKEK 730
DB 790 KSLRYLEENQETFPVLESNRPLRSLEVEEEOIRIVKLEKVSQDSLSGLAEENQPLR 849
QY 731 WVEEE-----RVFRGRADSFIAARMHLVQGRFRFSPWKGSVVDVSVIGVFL-----TONVS 780
DB 850 YLEEDDCINKSLLEDKTHKSIGLSLEDNRGDSIIP-----QESETOVSLRPPEEDQIV 904
QY 781 DHLSSAFMSLAARFPKPLSSSRREDENV--RSVVVEDPBGCIILNLNEIPSWQKVKOHP 838
DB 905 NHLEKES-----QEFSSRSSEEEQVMERSLEGENHE-----SLSSVEKEDQMVE--S 949
QY 839 DMEVSGVDG-----SKQLRDCNSNGIERFNFLEKSIQNL-----EE 876
DB 950 QLEKESQDSGKSLSEDESQETFGPLEKENAESLSLAGQDEEQKLEQETQOTLRAVNEQ 1009
QY 877 EVLSSQDSFDP-----AIFQSCGRVSGS--CSKSDAEFPFTR--CETKTVSQTS-- 922
DB 1010 MAVSPPEKVDPELPKPLGNQDEIARSLGKENQESLSLKEKGIETVKSLETEIIEPLETA 1069
QY 923 -----QSVOTGSPNLSDICLOGNERPHLYE-----GSGDVOKQETTNVAQKKPDLK 971
DB 1070 EEDLERKRSIDTOEPLWSTVARETVPEPDEPPGSLGSDENRETILTSLEKESQELSSL 1129
QY 972 MNW-----KDSVCFQGPNDTN-----WOTTPSSVYEQ----- 999
DB 1130 GKNNVTRVEDSQOCLQVEEGLOEHOESLRREVQBELPSSGNOQRWEDVVEGKAVGQEA 1189
QY 1000 -CATRQPHVLDIEDFGMQGEG-----LGYSWM--SISPRVDRVKNKNVPR 1041
DB 1190 PLATTGVGTEDKAEHLRGQGEAAAAEGELLQDIVGEAWSLGSSEPKQRPVPAEALDN 1249
QY 1042 RFRQGSVSPRETGQIIPSTPHLPOMGLSGSSSAVQEHODD--TOHQOQDEMKNASHLQ 1100
DB 1250 ---LEGAL-----EVP---VAQSMPEYTERDEDRQAAGEQDSIEVTGLGE 1289
QY 1101 --KTFDLNLSBECLTRQSSSTKQITDGLPRDRTAEDVVDPLSNNSLQNLIVSNSS 1158
DB 1290 AARTGLEL-----EQEVVGLDEPRHFAREBAIPPSLGESEVK----- 1326
QY 1159 NKEQTAVEYKETNATILREMKGTADGKPKTSQWDSLRKQVQEG--NEGRQERNKNMDSID 1217
DB 1327 -----AKIAQGLEG---PGKEP-----KEAGALDSGILELPKTSSEALE 1362
QY 1218 YEAIRRASISEIAKERGMNNMLAVRIKDFLERIVKHGGIDLEWLRSPPDKAKD-- 1275
DB 1363 CQHEE---SESMEGWEDEAS-----LE--TSDEHSGDAQOPR--PETEEDG 1405
QY 1276 YLLSIRGLKLSVECVRLTLHLNLAFPVDTNVGRIAVRMGWVPLQPLPESLQLHLLELYP 1335
DB 1406 AQAALTAQPKLLE-----FCSPILPILTAH--ELQP 1435
QY 1336 VLESIQKFLW 1345
DB 1436 QAECIQEAGW 1445
RESULT 7
ANK2_HUMAN STANDARD; PRT; 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).

Db 2336 AQESTATSDTKALPLPPEASVKTDTGTESKPGQVIRSPQGLLEALPSRDSVLSAVADDS 2395
Qy 291 LVSGNQQLGGPGNKRPIFLNHOITCLPAG-----NQLVGSPTDMHQLVMSTGGQOHGLLI 345
Db 2396 LA-----VSHKSLSPVLENSHKTDPDSL----- 2422
Qy 346 KNOQGSILIRQOQPCVPLIDQOATPKGFTHLNQMVATSMSSPGLRPHSQSQVPTTLHV 405
Db 2423 ---EPSPL--KESPCRDSSLESSVEPK-----MKAGIFP-SHPPLPAVAKT 2463
Qy 406 ESV-----SRILNGTGTQCRARAYDSLOD--IHQGNKYLISHET--SNNGCKKA 455
Db 2464 ELLTEVASVRSLRDLDPGSAED-----DSLEQTSMESSGKSPSPDTPSSSEVSVEY 2517
Qy 456 LPONSLPTPTIMAKLEA-----RGSKROYHRAMGTEKHDL-----NLAQQ 497
Db 2518 TPKTVDVSTPAVIEHCAEEDDSNGEKKR-----TPEEMFKVWTKMKMDELEQE 2571
Qy 498 IAQSQDV-----ERHNSST-----CVEYLDAAKTKIKQVQENLHGMPEVIEIEDDP 546
Db 2572 AKQRDYKKEPKQBESSSSDPDADCSVDVDEPKHTG---SGEDSGVPLVVTSESRKV 2627
Qy 547 TDGARKGNTASIGASKG-----NNSPVKTAKEKCIYVK- 584
Db 2628 SSSSESEPELAQLKKGADSGLLPEFVIRVQPPSLPSPMSDNSPPEYQFQP---VVSQ 2684
Qy 585 -----TPAKKGRAGRKKS-----VPPAHASEIQLWQPTTKPLRSRKP---GKGR 629
Db 2685 YTFKNEDTQEPKSEEEKSESHLAEDRAVSTEAEDRSYDKLNDRDTPQKICDHGC 2744
Qy 630 KSIQDSGAR-----GPSGELICQDSIAE--IYRNQNYLGDKEREQONAWLYKG 680
Db 2745 EAMSPSSARPVSSGLOSPTG---DDVDEQPVYKESIALQGTHEKDTGEEL----- 2794
Qy 681 DGAIVPYEKRRKPKPKVDIIDE---TTRINWLLMGKDEKGD----- 721
Db 2795 -----DVSRASPOADCSSEFSSSSSLPHCLVSEKDELDEISATSSIQTEVTKT 2846
Qy 722 EEDKKEKKEWEEERRVFRGRADF---IARMHLVQGDRRSPWKGSVVDS----- 769
Db 2847 DETFENLPKDCPSQSSITTTQDRFSDMDVPVSDLAENDEIYDPOITSPIYENVPQSFPSS 2906
Qy 770 -----VIGVFLTONVSDHL-----SSSAFMSLAARF----- 795
Db 2907 EESKTQTDANHTTFSHSEVYSVTITSPEDVVVAVSSSGTSLVSKESNFEGQDICKMSQL 2966
Qy 796 -----PPLKSSREDERNVRVWV-EDPEGICILNLEI--PSWOR-----KV 834
Db 2967 ESTLWEMQSDSVSSFPETMSGATTVVGEQISKVITTKTDVDSWSWSEIREDDAEFAEV 3026
Qy 835 QHPSDMEVSG--VDSGSK-----EQ-----LRDCSNSGI-- 861
Db 3027 KE-BEQKIFGLMVRQSGTTPDTPARTPEEGTPTSEQNPFQEGKLFEMTRSGAID 3085
Qy 862 -----ERNFL-----EKSIQNLSEEVLSQSDSPDPAIQSCGRVSGSCSKSDA--- 906
Db 3086 MTKRSYADESHFFQIGESREETLSEYDKEGATGADPLPLETSAESIALSESKETVDDE 3145
Qy 907 -----BFPTRCETKTVSGTSQS-----VOTGS-PNLS- 933
Db 3146 ADLLPDSVSEVEEIPASDAQLNSOMGISASTETPTKEAVSVGTKDLPTVGTGDIPLSG 3205
Qy 934 -DEICLOGNERP-----HLYEGSDVQKQETTNNVAQKKPDLKRTMMNKDSVC 979
Db 3206 VKQICSPDSSEPAVQVQLDFSLTRSVYSDRGD--DSPDSSPEEQSKSVIEIPTAPMENVP 3263
Qy 980 FGQPRN-----DTNQWTPSSSYEQCATQPHVLIEDFGMGGEGLYGSWMSISPRVD 1032
Db 3264 FTESKSKIPVTRMTPTSTAPPASAEYSSVS-----EDF-----LSSVDEENKAD 3307
Qy 1033 RVKNKN-----VPRFFRQ-----GGSVPREFT--GQIIPSTPHELPGMGLSGSSSAQVE 1080
Db 3308 EAKPKSKLPVKVPLQORVEQQLSDLDTSVQKTVAPQGGQDMASAPDNRSKSESASSLDSK 3367

RESULT 8

AIM1_HUMAN
ID AIM1_HUMAN STANDARD; PRT; 1723 AA.
AC Q9Y4K1; O00296;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Absent in melanoma l protein.
GN AIM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97250519; PubMed=9096375;
RA Ray M.E., Wistow G., Su Y.A., Meitzner P.S., Trent J.M.;
RT "AIM1, a novel non-lens member of the betagamma-crystallin
RT superfamily, is associated with the control of tumorigenicity in human
RT malignant melanoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234(1997).
CC -I- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
CC -I- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	ENBL; U83116; AAB53792.1; -;	
CC	ENBL; U83115; AAB53791.1; -;	
CC	HSSP; P02526; 1GAM.	
DR	MIM; 601797; -;	
DR	InterPro; IPR001064; Crystallin.	
DR	InterPro; IPR000772; Ricin_B_lectin.	
DR	Pfam; PF00030; crystall; 5.	
DR	Pfam; PF00652; Ricin_B_lectin; 5.	
DR	PRINTS; PR01367; BGCYSTALLIN.	
DR	SMART; SM00458; RICIN; 1.	
DR	SMART; SM00247; XTALBQ; 6.	
DR	PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 7.	
DR	PROSITE; PS00231; RICIN_B_LECTIN; 1.	
KW	Repeat; Lectin.	
FT	DOMAIN 1022 1119 CRYSTALLIN BETA/GAMMA-LIKE 1.	
FT	DOMAIN 1123 1206 CRYSTALLIN BETA/GAMMA-LIKE 2.	
FT	DOMAIN 1207 1318 CRYSTALLIN BETA/GAMMA-LIKE 3.	
FT	DOMAIN 1319 1404 CRYSTALLIN BETA/GAMMA-LIKE 4.	
FT	DOMAIN 1405 1497 CRYSTALLIN BETA/GAMMA-LIKE 5.	
FT	DOMAIN 1502 1584 CRYSTALLIN BETA/GAMMA-LIKE 6.	
FT	DOMAIN 1586 1719 RICIN B-TYPE LECTIN.	
SQ	SEQUENCE 1723 AA; 188646 MW; 7550F681A627FB09 CRC64;	
Query Match 2.3%; Score 207; DB 1; Length 1723;		
Best Local Similarity 18.7%; Pred. No. 0.0036;		
Matches 235; Conservative 170; Mismatches 475; Indels 376; Gaps 53;		
QY	18 QNDGSRDVLFEFLNKTPQOKPSKRBKPKVVEGKPKRKPAPLKVVEGPKR 77	
DB	91 ESDRSKO-----PPASPTKRGRSRALEAVAPPASGPRAPAK-----ESPKR 136	
QY	78 KPRKAQAEVKSKETGSAKKNLKESATKPPANVGDMNSKSPVTLKSKRKAL--NFDL 135	
DB	137 VPDPSPVTKTAESGEAARAIPRELVPKS-----SSILPEIKPKRKGPLPNHFN- 188	
QY	136 ENPDARQGDSESEIVONSSGANSFSEIRDAIGTNGSFLDSVQIDKTNGLGAMNOPLE 195	
DB	189 ---GRAEGGRSR-----ELGRAAGAPGASDADGLKPRNHF-GVGRSTVTTK 230	
QY	196 VSMGNQPKLSTGAKLARDQDPLLTRNQCQEPVATQNTQFPM---ENQAWLQMKNL 252	
DB	231 VTLPAKPKHVELNLKTPKNLD-SLGNENHNPFSQPVHKGNTATKISLFENKRT----- 281	
QY	253 IGFPFGNQPRMT-IRNQOQCLAMGNQOPMYLIGTPRPALVSGNQQLGGPGQGNKRP----- 307	
DB	282 -----NSSPRHTDIRPNTPASSKT-----FVGRKLNLAkakEMEQEKKVMPNSPQ 331	
QY	308 -----IFLNHQCLPAGNOLYSGPTDMHQL-----VNSTGG----- 338	
DB	332 NGVLVKETAETKTVSEELIPATRMNGDSSENQALGPQPNODDKADVQTDAGCLSEP 391	
QY	339 -----OOHGLLIKQOQCSLIRGQPCVPLIDQOPATPKGTHLNQVATSMSSPGL 390	
DB	392 VASALIPVKHKLKEDESEAA-----DSKSLVLENVTDTAQDIPITVDTKDL 439	
QY	391 RPHSQSQVPTTYLHVESVSRILNGTT-----GTCQRSRAPAYDSLQODIHQ 437	
DB	440 PPTAMPKQHTFDSQSPAESPPSLSAPAPGDVDPKTCVQSPISSEPCDTLKYSEN 499	
QY	438 NKYLISHEISNGCKKALPQNSLPTTPIA--KLEPARGSK-----RQVHRAMGQ 486	
DB	500 HKGCVL-PVSRQNNKMPLELGGTTPPLSTERSPEAVSECPSRVLVQVRSVLPVES 558	
QY	487 TEKHDNLNAQIQASQDVERHNSSTC-----VEYLDAAKTKIKVVOENLHG----- 534	
DB	559 TQ---DVS-SQVIPESSEVREVLPTCHSNPEPVVSVASCAPPQOEVLGNHSHCTAEALAA 615	

QY	535 -MPPEVIE-----IEDDPTGARGKGNKA----- 557	
DB	616 KSGPQVITPASEKTLPTAQOQSQRTPLAESSPTNSPSSGNHLATPQRDQTVTNGQDS 675	
QY	558 -----SISKASKG---NSSPVKKTAE-----KEKCIVPKTPAKKGRAGRK 595	
DB	676 PASLLNISAGSDSDSFSSDMKEFTIHKQMSAVCMKPKRKKARMPNSPA-----P 728	
QY	596 KSVPPPAHAHEIQLWQPTPKTPLSRSKPK---GKGRKSIQDSGKARGSGELLQDSI 651	
DB	729 HFAMPPIHEDHLE-----KVFDPKVFTEGLGKK--KES-----QPEM 763	
QY	652 AEIYRNQNLVLDGKREKQDQA--MVLYKGDGALVPYESSKKRPRKPKVDIDDETRINW 709	
DB	764 SPALHLMQNLDTKSLRPRKASAEQSVLEK---SLHTNTNGNSEPLVMPMEINDKENR--- 817	
QY	710 LLMGKGDEKDEKDKKKEKWEERVRGRADSFIAHMLVQGDRESPKWSVVD 769	
DB	818 -----DVTNGGIKRSRLE-----KSALFSSLLSLPODKIFSPSVTSVNTM 858	
QY	770 VIGVFLTONVSDHLSSAFMSLAARPPKLSRREDENRVSVVVEDEPGCILNLEIPS 829	
DB	859 TTAFTSTQNGS---LSQSSVSQPTTEGAPCGLNKE-----QSNLLPDNSLKVFNFSST 911	
QY	830 WQEKVQHPSDMEYSGVDGSKQELRDCNSGIERNFNFLEKSIONLEEVLSQDSQDPDAI 889	
DB	912 SHSLSKSPSHMEKVPQKEKTKEDLDSRSNLHLPETKFE--LSKLKNDMEKANHIESVI 969	
QY	890 FQSGRGVSCSCSKSDAE---FPTTRCEKTVSGTSGVQTSNLSDEICLOGNERPH 945	
DB	970 -----KSNLPNCANSDFDFMGLFKSSRYD-----PSIFSGLMSLSDTMTLRGVSQNK 1016	
QY	946 LYEGSGDVQKQETTNVAOKKPDLEKTMNWKDSVCFQGPNDNTNQTTPSSSYEOCATRQ 1005	
DB	1017 LNRPQKV-----VIYSEPDVS-----EKCI---E 1038	
QY	1006 HVLIDIEDFGMGEGLYGWSMISIPRDRVKNKNVPRFRFGGVSVPREFTQGIIPSTPHE 1065	
DB	1039 VFSIDIQCS-----SW-SLSP-VILIKVVRGCWILYEQ-----DNFEGHSIPLEGE 1083	
QY	1066 LPMGLSGSSSAVOEHQD-----DTHQNOODEMKNASHLOKTLFLLNLSSE 1111	
DB	1084 LELSLGLWIEDILIERHEEAESDKPVVIGSIRHVQD--YRVSH-----IDLFTPE 1132	
RESULT 9		
ZEPL_HUMAN	STANDARD; PRT; 2717 AA.	
ID	ZEPL_HUMAN	
AC	PI5822;	
DT	01-APR-1990 (Rel. 14, Created)	
DT	01-APR-1990 (Rel. 14, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-	
DE	binding protein 1) (HIV-Ep1) (Major histocompatibility complex binding	
DE	protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)	
DE	(PRDII-BF1).	
GN	HIVEP1 OR ZNF40.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=90169514; PubMed=2106471;	
RA	Fan C.M., Maniatis T.;	
RT	"A DNA-binding protein containing two widely separated zinc finger	
RT	motifs that recognize the same DNA sequence.";	
RL	Genes Dev. 4:29-42(1990).	
RN	[2]	
RP	STRUCTURE BY NMR OF 2113-2142.	
RX	MEDLINE=91064333; PubMed=2248949;	
RA	Omichinski J.G., Clore G.M., Appella E., Sakaguchi K.,	
RA	Gronenborn A.M.;	

KW		DNA-binding; DNA repair; Nuclear protein; Zinc-finger; Polymorphism.	
FT	ZN_FING	3042	3057
FT	ZN_FING	3086	3104
FT	VARIANT	231	231
FT	VARIANT	389	389
FT	VARIANT	1540	1540
FT	VARIANT	2607	2607
FT	CONFLICT	237	237
FT	CONFLICT	1156	1156
SQ	SEQUENCE	3130	AA; 352782 MW; 1C0700900F10BB14 CRC64;
Query Match		2.2%;	Score 200.5; DB 1; Length 3130;
Best Local Similarity		18.3%;	Pred. No. 0.019;
Matches		345;	Conservative 229; Mismatches 647; Indels 663; Gaps 88;
QY	6	DSSAVNATEAQEQDGRQDVLFLDNKTQQKPSKRKRKFKMPKVYVVEGKPKRKR---	61
Db	511	DNLSLLASLSLPQLDGTADENSDNPLNENSRTHS-----SVIATSKLSVKPSIFHK	562
QY	62	-----KPAELPKVVVEGKPKRKRKAATQEKVKSKE-----TGSAKKNLKSAT---	106
Db	563	DAATLEPSSAKITFOCK-----HTSALSHVLNKIEDLSQTNKNTKEGLDNSVTSF	617
QY	107	-----KKPANYGDM-----SNKSPVTLKCRKALNFDLEN--PGDARQGD--	146
Db	618	TNESTYSMKYPGSLSTSVHSENSHKESKKEILPVSCESI--FDYEDIPSVTRQVPSR	676
QY	147	-----ESEIV-----QNSGANSFSEIRDAIGTNGSFLDSVSDIKTNGLGA	189
Db	677	KYTNIRKIEKDSPTLHMHRHFNENTLGNPF-----NFDLNLHNSKNVS-----	720
QY	190	MNQPLEVSMGNQDKLSTGAKLARDQPDLLTRNQOQCFPVATONTOPFMENQQAWLQMK	249
Db	721	-----SEGNEKNSTALSFL-----PSSFTECELLSCSGENRTMVHSLNSTADESL	769
QY	250	NOLIGFPNGNOQPMRTIRNOQCLAMGNQOQPMYLIGTPRALVSGNOQLGGPQ-----	302
Db	770	NKL-KIRYEEQEHKT---EKPSLSQQAHYMFF-----PSVVLN-CLTRPQKLSPTY	819
QY	303	-----GNKRPIELNHQTLCPAGNQLYSGPTDMHQLVMSTGGQGHLLKNOOPGSLIRGQQ	358
Db	820	KLOPGNK-PSRLKLNKRLKLAGHQTSTKSS-----ETGSTKDNFIQNN-----	861
QY	359	PCVPLIDQOQATPKGFTHLNQMVATSMSSPGLRPHSQSOVPTT-----VLHVESV	408
Db	862	PCNSNPEKDNALASDLTKTT-----RGAFENKTPTDGFDICHFGDGTLETEQ-	908
QY	409	SRLILGTTGTQCRSRAPAY-----DSLQODIHQGNKYILSHEISNGNG-----CKKALPON	459
Db	909	SFGLVGNKYTLRAKKNVYETEDSESSFTVNTSKISLPHMEIGESLDGTLKSKRRKMS	968
QY	460	SSLPTPIAK---LEEARGSKROYHRAMQOTEKHD-----LNLAQQIAQSOQV---	504
Db	969	KKLP-PVLIKIYIIINRRGRKNMLVK-LGKIDSKEQVILTEEKMELYKKLAPLKDFWPK	1026
QY	505	-----ERHNSSTCEVLDRAKKT-KIQKVQENLHGMPEVIEIDDP	546
Db	1027	VPDSPATKIYIPLTPKSHRRKS--KHKSAKKTKGKOORTNENI-----	1070
QY	547	TGARKGNTAISKAGSKNSGSPVKKT--AEKEKCIVPKTPAKKGRAGRKSVPPPAHA	604
Db	1071	-----KRTLSFRKRSKSHAILSPSPSYNAETEDCDLNYSDVMSKLGFLSERSTSPINS	1123
QY	605	SEIQLWQTPPTPLRSKPKGKRKSIQDSGKARGPSGELLQCDSTAEIITYRMONLY--	662
Db	1124	SPRCWSPTDPR-----AEEIMAAAEKAMLFKGNVYKK	1158
QY	663	-----LGDKEREQEQNAMVLYKGDGALV--PYESKKRPRPKVIDDETTRINWLLMGKG	715

Db	1159	TVNSRIGKTSRAQAQ-----IKKSKAKLANPSIVTKRNRK-----NOTNKLVD-----	1202
QY	716	DEKEGDEEKKKKEKWKWEEERRVFRGRADSFITARMHLVQGDRRFSPWKSGVSDVSVIGVFL	775
Db	1203	DGKKKPRAKOKTNEKG-----TSRKHITLTKDEKIKSQSGAEVAFVL---K	1244
QY	776	TQNVSDHLSSSAFMSLAARFPPKLSRSSREDNRVRSVVVEDPEGCILNLNEIPSWQEKVQ	835
Db	1245	HQNVSEFASSSGGSQSL-----LPKQKDFMLMSAVD	1275
QY	836	HPSDMEV-SGVDSGSKQLRDCNSGIERENFLEKSIQNLEEEVLSSQDSFDPALFQSCG	894
Db	1276	HPLSASLPTGIN-AQKLSGCSFSSLES---KKSVD--LQTFPSSRDDLHFSV--CN	1325
QY	895	RVGSCSCSDAEFPPTTRC-----ETKTVSGTSSQSVQT	927
Db	1326	SIGP-GVSKINVQRPHNQSAMFTLKESTLIQKNIFDLNSHLSQVAQNTQISSGSSKIED	1384
QY	928	GSPNLSDEICLOGNERPHLYEGSGDVQKQETTVNAQKKPDLEKTMW---KDS---VCF	980
Db	1385	NANNI-----QRNYL---SSIGKLSYRNSLESKLDQAYTPNFKHCKDQQQIVCI	1432
QY	981	GQ-----PRNDTN-----WQTPSSSYEQCATRQPH---VLDIEDFGMQEG	1019
Db	1433	AEQSKHSETCSPGNATASEESQMPNCFVTSLRSPKIQIAWEQKQRGFILDMSNFKPE---	1489
QY	1020	LGYSWMSISPR-----VDRVKNVKNVPRFRFGQGSVPREFTGQIIPSTPHELPG	1068
Db	1490	-----RVKPRSLSEALSQTKALSQCKNRV-----STPSAF-----G	1521
QY	1069	MGLSGSSSAQVEHQDDTQHNOQDEMKNASHLQKTLFDLLANSSECLTROSSTKQNTDGC	1128
Db	1522	EGOSG-----LAVLKELQKROOKAON-----	1543
QY	1129	LPRDRTAEDVVVDPLSN---NSSLQNILVESNSNKBQTAVEYKETNATILREMKGTAD	1184
Db	1544	-----ANTQDPLSNKHQPNKNSIGSL-EHNKANKRTRSV-----TS	1579
QY	1185	GKPTQOWDSLKRDVEGNEGRQERNKNMDSIDYEATRASISIEISAIKERGMNNMLAV	1244
Db	1580	PRKPTPRSTKQKE-----KIPKLLKVDSLN---L	1606
QY	1245	RKDFLERIKVHDGIDLEWLRESPPDKADYLLSIRGLGLKSVCEVRLTLHLNAPVD	1304
Db	1607	QNSQLDNSVSDSPFPF-----SDPGFESCYSLE-----DSLSPHNYNEDIN	1650
QY	1305	T--NVGRIAVRMG--WVPL-QPLPES-LQLHLLELYPVLESIOK--FL-----	1344
Db	1651	TIGQTGFCFSYSGSQFVPADQNLQKFLSDAVQDLFPG-QAIEKNEFLSHDNQKCEDKH	1709
QY	1345	-----WPLRCKLD-----QRTLYELHYQLITFGVKVCTKSRPN	1377
Db	1710	HTTDSASWIRSGTSLSPSEIFKSTIDSNEENRRHWNKNSFHPLTTRSNSIMDSFCVQQAED	1769
QY	1378	CNACPMRGECHRFASASARLALPAPEE-----RSLTSAT-----IPVPP	1418
Db	1770	CLSEKSR---LNRSSVSKEVFLSLPQPNNSDWIOGHTRKEMGQSLDSANTSFATLLSPD	1826
QY	1419	ESFPPVAIPMTEL-----PLPLEKSLASGAPNSREN-----CEPIIEPA	1458
Db	1827	GELVDVACEDLELYSVRRNDMLTPTPOSSPRSTSSPSQSKNGSFTPTRTANILKPLM-SPP	1885
QY	1459	SPGECETEITSDIEDAYNE-----DPDEITKLNITBQFGMTLREHMERNEWLEQGDMS	1514
Db	1886	SREEIMATLLDHLSETIYQBPFCNSPNSDVPEKPREIGRLLMVETRLANDLAIEFEGDFS	1945
QY	1515	KALVALHPTTTSITPPLKNISRLRTEHQVVELPDSHRLLDGMDKRE-----PDDPSPYL	1569
Db	1946	LEGURLWATASMTQNPGRPSGOGVNVKNGSSNPKMWVEDKKIVIMPCCKAPSRQL	2005
QY	1570	LAIWTPG--ETANSAQPPQKCGG	1591

Db 2006 VOVWLOAKEEYERSKKLPKTPGTG 2029

RESULT 11
AKAC_HUMAN
ID AKAC_HUMAN STANDARD; PRT; 1781 AA.
DC 002952; Q99970; Q00498; Q00310;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE A-kinase anchor protein 12 (A-kinase anchor protein 250 kDa) (AKAP
DE 250) (Myasthenia gravis autoantigen gravin).
CN AKAP12 OR AKAP250.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RN
RC TISSUE=Heart;
RX MEDLINE=97153077; PubMed=9000000;
RA Nauert J.B., Klauck T.M., Langeberg L.K., Scott J.D.;
RT "Gravin, an autoantigen recognized by serum from myasthenia gravis
RT patients, is a kinase scaffold protein.";
RL Curr. Biol. 7:52-62(1997).
[2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=98269042; PubMed=9604001;
RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT "Changes of gene expression by lysophosphatidylcholine in vascular
RT endothelial cells: 12 up-regulated distinct genes including 5 cell
RT growth-related, 3 thrombosis-related, and 4 others.";
RL J. Biochem. 123:1119-1126(1998).
[3]
RP SEQUENCE OF 43-1781 FROM N.A.
RN
RC TISSUE=Umbilical vein endothelial cells;
RA Bowditch R.D., Ginsberg M.H.;
RT "Sequence of gravin cDNA isolated from a human umbilical vein
RT endothelial cell library.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 1477-1781 FROM N.A.
RN
RC TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=92395179; PubMed=1522245;
RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,
RA Lindstrom J., Ginsberg M.H.;
RT "Molecular cloning and preliminary characterization of a novel
RT cytoplasmic antigen recognized by myasthenia gravis sera.";
RL J. Clin. Invest. 90:992-999(1992).
CC -!- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR
CC (PKC).
CC -!- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, LEUKOCYTES,
CC FIBROBLASTS AND OSTEOCARCOMA. BUT NOT IN PLATELETS, LYMPHOCYTES,
CC MONOCYTIC CELL LINES OR PERIPHERAL BLOOD CELLS.
CC -!- INDUCTION: ACTIVATED BY LYSPHOSPHATIDYLCHOLINE (LYSOPC).
CC -!- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE
CC INVOLVED IN BINDING PKC.
CC -!- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
CC PATIENTS WITH MYASTHENIA GRAVIS (MG).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch).

CC EMBL; U81607; AAC51366.1; -.
CC DR EMBL; AF001504; AAB58938.1; -.
CC DR EMBL; AB003476; BAA19927.1; -.
CC DR EMBL; M96322; AAA35931.1; -.
CC DR MIR; A43922; AAA3922.1;
CC DR PIR; 604698; -.
CC DR InterPro; IPR001573; PkinA_anch.
KW Antigen; Repeat; Alternative splicing.
FT DOMAIN 603 633
FT AKAP 1.
FT DOMAIN 752 782
FT AKAP 2.
FT DOMAIN 797 827
FT AKAP 3.
FT DOMAIN 98 101
FT POLY-GLU.
FT DOMAIN 265 556
FT INVOLVED IN PKC-BINDING (PROBABLE).
FT DOMAIN 1540 1553
FT RII-BINDING (PROBABLE).
FT VARSPLIC 1 98
FT MISSING (IN ISOFORM 2).
FT VARSPLIC 99 106
FT EEEVIVTE -> MLGTTTIT (IN ISOFORM 2).
FT CONFLICT 117 117
FT E -> K (IN REF. 2).
FT CONFLICT 142 144
FT NRN -> TPEI (IN REF. 2 AND 3).
FT CONFLICT 215 215
FT Q -> K (IN REF. 2).
FT CONFLICT 448 448
FT G -> E (IN REF. 2 AND 3).
FT CONFLICT 694 694
FT R -> G (IN REF. 2 AND 3).
FT CONFLICT 867 867
FT S -> S (IN REF. 2 AND 3).
FT CONFLICT 986 986
FT S -> A (IN REF. 2 AND 3).
FT CONFLICT 1530 1530
FT E -> EE (IN REF. 3 AND 4).
FT CONFLICT 1581 1581
FT V -> M (IN REF. 4).
FT CONFLICT 1601 1601
FT Q -> L (IN REF. 2).
SQ SEQUENCE 1781 AA; 191439 MW; BA813937379FAC0F CRC64;

Query Match 2.1%; Score 192.5; DB 1; Length 1781;
Best Local Similarity 17.7%; Pred. No. 0.023;
Matches 276; Conservative 217; Mismatches 567; Indels 499; Gaps 72;

Qy 105 ATKRPANVGDMSNKSPEVTLKSCRKALNFDLENPGDARQDQSESEIVONS-----SCAN 158
Db 23 AEPEPGSGGSAEAAAPDTT-----ADPAIA-ASDPATKLLQKNGQLSTINGVA 69
Qy 159 SFSEIRDAIGGTNGSFLDSVSQIDKTNGLGAMNPLE-----VSMGNQPD 203
Db 70 EQDELSQLQEGDLNG-----QKGNALNGQALNSQEESEVIVTEVGQRSDVSESDSK 122
Qy 204 KLSTGAKLARDQOPDLLTRNQOC-QFPVATQN-----TQPPMENOQAWLOMKN--OLIGFP 256
Db 123 EMATKSAAVVDITDDGQENRNIEQIPSSSENLBELTQ-PTESQANDIGFKKFKVGF 181
Qy 257 FGNQOQPR-----MTIR-----NQOPCLAMG-----NQOPMYLIGTPRPA 290
Db 182 FTVKKDKTEKPTDVTOLLTVKKDEGEAGAGADHDQDPSLGAGEAASKESEPKQ--STEKPE 239
Qy 291 LVSGNQQLGGPGQGNRPFLFNHQTCLPAGNQLYSPTDMHOLVMSTGGQOQHLLIKNOOP 350
Db 240 ETLKREQ-----SHAEISP-----PAESQAVEEC--KEGEEKQKEP 276
Qy 351 GSLIRGOQPCVPLIDQOPATPKGFTHLNQMVATSMSSPLRPHSQSQVPTTYLHVESYR 410
Db 277 SK--SAESPTSPVTSETGTSTFKKF--FTQGWAGWRKKTSEFKPKDEVEAS----- 323
Qy 411 ILNGTTGTCQRAPAYDSLOODTHQGNKYILSHIEISNGNGCKKALPONSSLPTPIMAKL 470
Db 324 -----EKKKEQEPKVDTEED---GRAEVASEKLTASEQAHPQEPAESAEHPRLSAEY 373
Qy 471 EEAR-GSKROYHRAWGOTEKHDNLNLAOOIAQSDOVERHNS-----TCVEYDLAAKKT 522
Db 374 EKVELPSEEGVSGQGPSEKPAPLATEVF-DEKIEVHQEEVVAEVHVSTVEETEEDQ 432
Qy 523 KIQK-----VVOENLHGMPEVIEIE-----DDPTDGARKGNKTASISKGA 563
Db 433 EVEETAGSVPAEELVGMDAEPQAEPAKELVKLCTCVSGEDPTQGDLSDEKVLSP-P 491
Qy 564 SKGNSSPVKKTAEEKICIVPKTPAK-----KGRAGRKSVPPPAHASEIQLWQPTPP 615

Db 492 PEGVSEVMSLQERMKVGSPLKLTFTGLKLGKQKQKRGSGDGESEHTQVPA 551
Qy 616 KTLPSRSPKPKGKSTODSGKARGPSGELL- QDSTAEIIRMQNLYLGDKEREQDNA 674
Db 552 DSPDSQEQKGESSASPEPE-----EITCLEKGLAEV-----QDGEAEEGA 595
Qy 675 MVLYKGDG-----ALVPYSEKKRPRKVDIDDETTRINLLMKGDKDEKDEKKKE 729
Db 596 ----TSDGKKREGVTPWASFKKMVTPKKRV-----RRPSEKDELDKVK- 638
Qy 730 KWEEERRVFRGRADSTFIARHLVQGRDRFPWKGVSDVSVIGVFLTQNVSDHLSAFA 789
Db 639 -----SATLSSTF----- 646
Qy 790 SLAAREFPKLSRSDERNRVSVVVEDEPGCILNINLIPSMQEKV-----QHPS 838
Db 647 STASEMEEMKGSVEEPK-----PEPKRKVDTSVSEALICVSGSKKRRARRSS 697
Qy 839 DMEVSGVDGSGKEQLRDCNSGIERFNFLKSIQNLSEELVSSODSFDPAIFOSCGRVGS 898
Db 698 DEEGPKAMGDHQKAD-EAG-----KDKETGTDGILAGSQEHDP-----GQ 738
Qy 899 CSCSKSDAEPTTRCETKTVSGTSQSVQTSQSPNLSDEICLQGNRPHLYEGSGDVQKQET 958
Db 739 GSSPEQAGSPT-----EGEGSVTWSFRLVTPRKKSLEEKSEDSI-ACSG----- 787
Qy 959 TNVAKKPDLEKTMNWKDSVCFGQPRNDTNW-----QTPSSSYEQCATRQPHV 1007
Db 788 --VEHSPDTE-----PGKEESVSIKKFIPGRKKRPDKGQEA----- 825
Qy 1008 LDIEDFGMGGLGYSM-STSP--RVDVRKNKNVPRFRFGGGSVPR-----EFTQII 1059
Db 826 -PVEDAGTGANEDSDVPVAVPLSEYDAVEREKNEAQQAQKGAQEQPEAKAATEVSKELS 884
Qy 1060 PSTHELPGMGLSGSSA--VOEHQDDTQHQQDEMKNASHLQKTFDLNLSSECLTRQ 1117
Db 885 EQVHMAAAVADGTRAATITERSPSW-----ISASVTEPLEQV 924
Qy 1118 SSTQNITDGLPRDRAED-----VVDPLSNSSLQ--NILVESNKNKEQTAVEYKETN 1171
Db 925 EAAALLTVEVLREVIAREEPTVTEPLPENREARGDTVVSEALTPAEVATAAE---T 980
Qy 1172 ATILRMKGTLADGKKPTSQWDSL-----RKDVEGNEGROERKNKNMDS 1215
Db 981 APLGSEEGTASAAEETEMVSAVSLQDSDPTTEATPVOEEGVPDIEEQERRTQE 1040
Qy 1216 IDYEAIRRASISSEATKER-----GMNMLAVRIKDLERIVKDHGIDLEWLRESPPD 1271
Db 1041 V-----LQVNAEKVESQLPGTGP-----EDVLQPVQRAEA-----ERPEE 1078
Qy 1272 KAKDYLLSIRGLGKSVCEVRLTLHNLAFVVDNNGRI-AVRMGWVPLQPLPESLQHL 1330
Db 1079 QAE-----ASGLKKE-----DVVLKVDQAQEAKEPTFGKVVQQTTPESFE--- 1120
Qy 1331 LELYPVLESIOKFLWPLCKLDQRTLYELHYQLITFGVKTRKSRPNCNACPMRGECHRF 1390
Db 1121 -KAPQVTESIES-----SELVT-----TCQAEFLAG----- 1145
Qy 1391 ASAYASAKLALPAPERSLTSATIPVPSFPPVPAIPMIELPLP-----LEKS 1438
Db 1146 ---VKSQEMV-----EQAIIPDSVETPTDSETDGTSPVADFDAPCTOKDELVEIHEENE 1198
Qy 1439 LASGAPSNRENCEPI---IEEPASPG-----QECTEITESIEDAYNEDPD-EIPTIKL 1489
Db 1199 VASGTQSGTEAEAVPAQKERPPAPSSVFQEBETR-EQSKMEDTLHTDKVSVETVSI 1256
RESULT 12
ID BRCL_MOUSE STANDARD; PRT; 1812 AA.
AC P48754; Q60957; Q60983;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Breast cancer type 1 susceptibility protein homolog.
GN BRCA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96177659; PubMed=8634697;
RA Abel K.J., Xy J., Yin G.Y., Lyons R.H., Meisler M.H., Weber B.L.;
RT "Mouse Brca1: localization sequence analysis and identification of
RT evolutionarily conserved domains.";
RL Hum. Mol. Genet. 4:2265-2273(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=96177660; PubMed=8634698;
RA Sharan S.K., Wims M., Bradley A.;
RT "Murine Brca1: sequence and significance for human missense
RT mutations.";
RL Hum. Mol. Genet. 4:2275-2278(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96121367; PubMed=8575748;
RA Bennett L.M., Haugen-Strano A., Cochran C., Brownlee H.A.,
RA Fiedorek F.T. Jr., Wiseman R.W.;
RT "Isolation of the mouse homologue of BRCA1 and genetic mapping to
RT mouse chromosome 11.";
RL Genomics 29:576-581(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Embryo;
RX MEDLINE=96067162; PubMed=7590247;
RA Lane T.F., Deng C., Elson A., Lyu M.S., Kozak C.A., Leder P.;
RT "Expression of Brca1 is associated with terminal differentiation of
RT ectodermally and mesodermally derived tissues in mice.";
RL Genes Dev. 9:2712-2722(1995).
RN [5]
RP SEQUENCE OF 727-1111 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96021028; PubMed=7550308;
RA Marquis S.T., Rajan J.V., Wynshaw-Boris A., Xu J., Yin G.Y.,
RA Abel K.J., Weber B.L., Chodosh L.A.;
RT "The developmental pattern of Brca1 expression implies a role in
RT differentiation of the breast and other tissues.";
RL Nat. Genet. 11:17-26(1995).
RN [6]
RP SEQUENCE OF 789-1250 FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96163506; PubMed=8566965;
RA Schroeck E., Badger P., Larson D., Wynshaw-Boris A.,
RA Ried T., Brody L.;
RT "The murine homolog of the human breast and ovarian cancer
RT susceptibility gene Brca1 maps to mouse chromosome 11D.";
RL Hum. Genet. 97:256-259(1996).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN
CC TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY
CC SIMILARITY).
CC -!- SUBUNIT: CTIP INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN OTIC VESICLES AT
CC DAY 9.5. AT DAY 10.5, THIS EXPRESSION DECREASES AND HIGH LEVELS
CC ARE FOUND IN THE NEUROECTODERM. AT DAYS 11-12.5, HIGH LEVELS IN
CC DIFFERENTIATING KERATINOCYTES AND WHISKER PAD PRIMORDIA. AT DAYS
CC 14-17, EXPRESSION ALSO OBSERVED IN KIDNEY EPITHELIAL CELLS. IN
CC THE ADULT, HIGHEST LEVELS FOUND IN SPLEEN, THYMUS, LYMPH NODES,
CC EPITHELIAL ORGANS, AND ALVEOLAR AND DUCTAL EPITHELIAL CELLS OF
CC THE MAMMARY GLAND. VERY LOW LEVELS IN BRAIN, KIDNEY, AND SKIN. NO
CC EXPRESSION IN HEART, LIVER OR LUNG.

Qy 1047 GGSVPREFTGIIIPSTHELPMGLSGSSAVQRBHDDTQHNOODEMKASHLQKTFIDL 1106
 Db 1440 GPTSGDDSGMRFS-PFKSPLAGRSAGHCCSRHLQKRNSPSEELLQAPAG----- 1490
 Qy 1107 LNSSEECITROSSTKONIT-DCGLPRDRTADVDVPLNNSSLQNLIVLSSNNKEQTAV 1165
 Db 1491 -----SEASSEPHNSTGSCPLR-RELEGP-TPYLGSGLFSSRDPSESPKEPAHI 1540
 Qy 1166 EYKTNATILREMGKTLADGKKPTSQWD-----SLRKDVNEGREGQERKNMDS-- 1215
 Db 1541 GTTTPASTSVLRKPOQVAFRAAAAGADKAVGVIVSKIKPELTSEERADRIDSMVYSGL 1600
 Qy 1216 IDYEAIRRASISE-----ISEAIKRGNNMLAVRIKDFLERIVKDHGGIDLEWLSRPP 1270
 Db 1601 TPKEVMTQKFAEYKRLTLDAITEETHVIKTDAEVCERTLK----- 1645
 Qy 1271 DKADYLLSIRG-----LGLKSEVCVRLTLHLNLAFFVDVTNVR 1309
 Db 1646 -----YFLGIAGGKWIVSYSWVRSIQERLLNHEFEVTDGVDVTGR 1687

RESULT 13
 CBP_HUMAN
 ID CBP_HUMAN STANDARD; PRT; 2442 AA.
 AC Q92793; Q16376; O00147;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CREB-binding protein.
 GN CREBBP OR CBP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=97385172; PubMed=9238046;
 RA Sobulo O.M., Borow J., Tomek R., Reshimi S., Harden A.,
 RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
 RA Zeleznik-Le N.J.;
 RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
 RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
 [2]
 SEQUENCE FROM N.A.
 MEDLINE=97321049; PubMed=9177780;
 RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
 RA Lushnikova T., van Omnen G.J.B., Goodman R.H., Deaven L.L.,
 RA Doggett N.A., Peters D.J.M., Breuning M.H.;
 RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
 RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
 RT 16p13.3";
 RL Genomics 42:96-144(1997).
 [3]
 SEQUENCE OF 1-405 FROM N.A.
 MEDLINE=96376968; PubMed=8782817;
 RA Borow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,
 RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
 RA Horman D., Mittleman F., Volinia S., Watmore A.E., Housman D.E.;
 RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
 RT a putative acetyltransferase to the CREB-binding protein.";
 RL Nat. Genet. 14:33-41(1996).
 CC -1- FUNCTION: MEDIATES GENE REGULATION BY BINDING SPECIFICALLY TO
 CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
 CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CC CAMP-RESPONSIVE GENES.
 CC -1- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(8;16)(p11;p13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
 CC INVOLVING CBP AND MLL.
 CC -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI

CC SYNDROME (RTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
 CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
 CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; UA7741; AAC51770.1; -;
 CC EMBL; U85962; AAC51331.1; -;
 CC EMBL; U89354; AAC51339.1; -;
 CC EMBL; U89355; AAC51340.1; -;
 CC MIM; 600140; -;
 CC InterPro: IPR001487; Bromodomain.
 CC InterPro: IPR003101; KIX
 CC InterPro: IPR000197; TAZ_finger.
 CC InterPro: IPR000433; ZnF_ZZ.
 CC Pfam; PF00439; bromodomain; 1.
 CC Pfam; PF02135; KIX; 1.
 CC Pfam; PF02172; zf-TAZ; 2.
 CC Pfam; PF00569; ZZ; 1.
 CC PRINTS; PR00503; BROMODOMAIN.
 CC SMART; SM00297; BROMO; 1.
 CC PROSITE; PS00633; BROMODOMAIN_1; 1.
 CC PROSITE; PS00014; BROMODOMAIN_2; 1.
 CC PROSITE; PS01357; ZF_ZZ_1; 1.
 CC PROSITE; PS01355; ZF_ZZ_2; 1.
 CC Transcription regulation; Nuclear protein; Activator; Bromodomain;
 KW Chromosomal translocation; Zinc-finger.
 FT ZN_FING 1701 1744
 FT DOMAIN 363 430 CYS/HIS-RICH.
 FT DOMAIN 452 683 CREB-BINDING.
 FT DOMAIN 1103 1175 BROMODOMAIN.
 FT DOMAIN 1061 1064 POLY-GLU.
 FT DOMAIN 1199 1487 CYS/HIS-RICH.
 FT DOMAIN 1555 1562 POLY-GLU.
 FT DOMAIN 1675 1849 CYS/HIS-RICH.
 FT DOMAIN 1943 1948 POLY-PRO.
 FT DOMAIN 1967 1970 POLY-GLN.
 FT DOMAIN 2081 2085 POLY-GLN.
 FT DOMAIN 2199 2216 POLY-GLN.
 FT DOMAIN 2245 2248 POLY-GLN.
 FT DOMAIN 2297 2300 POLY-GLN.
 FT CONFLICT 1511 1513 FAE -> NSG (IN REF. 2).
 FT CONFLICT 1724 1725 ED -> VV (IN REF. 2).
 FT CONFLICT 1770 1770 V -> L (IN REF. 2).
 FT CONFLICT 1789 1789 N -> F (IN REF. 2).
 FT CONFLICT 1812 1812 T -> P (IN REF. 2).
 SQ SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;
 Query Match 2.1%; Score 192; DB 1; Length 2442;
 Best Local Similarity 18.1%; Pred. No. 0.038;
 Matches 345; Conservative 238; Mismatches 720; Indels 598; Gaps 89;
 QY 5 MDSSAVNATATEQDNGSRQDVLEFDLNTK-----POQPSKRRKRPMPKVVEGKPKRK 59
 Db 288 MGATGVNPQLASKQSMVNSLPTFTPTDIKNTSVTNVPMNSQMOTSVGVPTQATATGPTAD 347
 QY 60 P--RKPAELPKVVVEGPKPKPKAAQTKVSKETGSAKKKLNKESATK----KPAVNG 113
 Db 348 PEKRLQLQQLVLLHLLHAKCORREQANGEVACSLPHCRPTMKVNLNHTHCQACKAQA 407
 QY 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDARQGD----SESEIVQNSGANSFSFIRD 165
 Db 408 HCASSRQIISHWNKTRHDCPVCPLPLKNASDKRKNQQTILGSPASGIQNTIGS-----V 460

Qy	166	AIGTNGSF	LD	SVSQ	IDKYN	-----GLGAMNOPLEV	-----	196
Db	461	GTGQONAT	SL	SNPN	IDPSSMGR	AVAAALGLPYMNO	QOTOLQPVGPQO	PAQPTHQOMRT
Qy	197	--SNGNOP	DL	STGAK	LARDQO	PDLLTRNQO	QCFPVATONTQ	FPWENQOAWLQMNOLIG
Db	521	LNPLGNPN	MI	PAGG	-ITTDQPPN	LISEALPTSL	GATN	---PLMND
Qy	255	FPFGNQPR	MTIR	NQOPL	AMGNOQPMV	-----	LIGTPPA	-----
Db	566	SNSNGITL	STPTA	PPSTG	VRGKGW	HEHVTDRLSH	VHLVQA	ITPTDPAALKDRR
Qy	291	---LVSGN	OOLG	---POGN	KRPIFLNH	QTCPLPAGN	OLYSGPTDMH	OLVMSGTGQOHHLL
Db	626	MENLVAY	AKKVEG	DYME	SANSRDEYVH	-----LLAE	KIYKIQKELE	KRRS---RLHKQG
Qy	345	IKNQPGSL	IRGO	QCPV	LIDQ	-OPATPKG	FTHLQMWAT	SMSSPGLRPHS
Db	678	ILGNQPAL	PAPGAQ	P	-PVIQAQ	VPRP	-----	PNGPLSLPVNRM
Qy	404	HVESVSR	ILNG	-TTGTQ	CRSRAPAY	DSLQDDIH	QGNKVILSHE	ISNGNGCKKALPONSSL
Db	716	QVSGHNS	FNPN	SLGN	VQLPQAPM	-----	GPRASP	MNHSV
Qy	463	PTPIMAK	LEEAR	GSRK	QYHRAM	GOTEKHDLN	-LAQOIA	QSQDVERHNS
Db	752	QMSMGSV	PCMA	ISPR	MRPQPPNM	GAHTNNMA	QAPASQFL	PQNQFPSSSGAMS
Qy	519	-----	GMPPE	IEID	PTD	GARKGNT	ASISKG	---ASKGNSP
Db	812	OPPAQT	GVSG	QVGP	GAALPN	LMGLPQAS	QPCPV	QTSPLHTPPPA
Qy	534	---GMPPE	IEID	PTD	GARKGNT	ASISKG	---ASKGNSP	-YKKTAEKECIVPK
Db	872	TTTPGHT	PPQPA	PTQSP	TPVSSSG	QTPPTP	PGSVPS	ATQTSPTVQA
Qy	585	TPAKKGR	AGRRKS	---VPPAH	-----	ASEIQL	WQTP	-----PK
Db	932	TPVQPS	VATPQSS	QOQPT	PVHAQ	PGPTPLS	QAAASIDNR	VPTSSVASAETNS
Qy	617	TPLSRK	PKGGRKS	IQD	SGKARG	-PSGELL	CODSIAELI	YRMQNL
Db	992	VPLEKMT	ETAQ	EDT	PD	PGESK	GP	PRSMEE
Qy	676	VLKGDG	ALVP	YESK	RRKRP	KVDIDDET	TRIWNLL	MGKDEKEDK
Db	1039	AEQKSE	-----	PMEV	DEKKEV	KVEVKEE	BESSN	---GTASQ
Qy	736	RRVFRGR	ADSF	IAR	MHLV	QDGR	RRFPW	KSGSVVDSY
Db	1091	QOALMPT	LEA	-----	LYROD	PESLP	FRQ	PDPLLGI
Qy	796	PKLLSS	REDE	-RNRV	SVVVED	PEGCIILN	-LNEIP	SWKEKVQHP
Db	1141	--KLD	GOYQ	EPWQ	YVDDY	WLMFN	AWLN	NRKTSVYK
Qy	846	DSGSK	-----	EOL	---RDC	NSGTE	-RFNFLE	KSIQNL
Db	1199	CCGRKE	FEP	QTLCCY	GKQCL	TI	PRDAAY	SVQNRHY
Qy	881	-----	SQDS	FDPA	IFQSG	GRVGS	-----	C-SCKS
Db	1259	QTTISK	DOPE	KKND	TLDP	PEP	VDC	KECGRK
Qy	909	PTTRCET	KTV	SGTS	QVQ	SGSNL	SD	EI--CLOG
Db	1318	--PRKEN	KFS	AKR</				

Qy	1021	GYSWMSISPRVDRVKNKNVPR--RFRFGQGVSPR-----EF-----TGQ	1051
Dy	1021		1051
Db	1418	GSD--CPPPNTRRVISYLDLSIHFR-----PCLRTAVYHEILIGLEYVKKLGYYVTGH	1470
Qy	1058	IIPSTPHELPGMGLSGSSSAVQEHQDDTOHNOODEMKASHLQKTFDLDLNS--SEECLT	1115
Dy	1058		1115
Db	1471	IWACPPSE-----GDGYIFCHPPD-----QKIPKPKRLQEWYKMKLDKFAERI	1517
Qy	1116	RQSTKQNIITDGLCPDRDAEDVDPPLSNNSLOWILVSSNNKEQTAVYKTKTNATIL	1175
Dy	1116		1175
Db	1518	DYKIDFQKATÉ-----DRLTSAKELPYFGDPWPVNVLESIKELEQEEERKKESSTAAS	1572
Qy	1176	REMGKTLIADGKPKTSQWDSLRLKDVEGNEGRQERNKNNMDSIDYEAIRRASISEIAIKE	1235
Dy	1176		1235
Db	1573	ETTEGSGDSK-----NAKKKNKKTKNKK-----SSISRANKK	1606
Qy	1236	R---GMNNLAVRIKDFLER-----IVKDHGGIDLEWLRESPPDKAKDYLLSIRGLGL	1285
Dy	1236		1285
Db	1607	KSPMPNVSNDLSQKLYATMEKHKEVFFVILHLAGVINTL---PPIVDPPDLLS-----	1657
Qy	1286	KSVECVRLLTLLHNAFFVDITNVGR-----IAYRWGVPQLPLPESQLHLELYPVLESI	1340
Dy	1286		1340
Db	1658	-----CDLMDGRDAFLTARDKH-----EFSLL	1681
Qy	1341	QKFLWPRCLKDQRTLYELHYQLITFGK---VFCTKSRPNCNACPMRGECRHFASAYASA	1397
Dy	1341		1397
Db	1682	RRSKWSTLC---MLVELHTQ---GQDREVV-----TCNECKHHVETRMHCTVCEDY	1726
Qy	1398	RLALPAPEERLSINATIPVPPSPFPVAIPMTELPL-EXSLASGAPSNRENCEPIIEE	1456
Dy	1398		1456
Db	1727	DLCTCNCYNTRS-----HAHKMYKWLGLDDEGSGQGEPSK-----	1762
Qy	1457	PASPGQCEIT-----ESDIEDAYYNEDDPIPTIKLNIQFGMTLREHMERNMELQEG	1511
Dy	1457		1511
Db	1763	--SP-QESRRVSTORCIQSLVHACQCRNANCSLPS-----QMKRVVQHTKG	1807
Qy	1512	-----DMSKALVAL-----HPTTISPTPKLKNIS-RLR---TEHOVYELPDSDHR	1552
Dy	1512		1552
Db	1808	CKRKTTNGCGPVCKQIALCYHAKHKQCNKCPVPCLNKKHKLRQQQIQHRLQQAQLMRR	1867
Qy	1553	LLDGMQDKR---EPDPSPYLLAIWTPGE---TANSAQRPQEQ	1587
Dy	1553		1587
Db	1868	RMATMTNTRNYPQOSLPSPTSAPPGTPTQOPSTPTQPPQPPAQ	1908
Qy	1587		1908
RESULT	14		
TPR_HUMAN			
ID	TPR_HUMAN	STANDARD;	PRT; 2349 AA.
AC	P12270;		
DT	01-OCT-1989 (Rel. 12, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Nucleoprotein TPR.		
GN	TPR.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID	9606;		
RN	[1]		
SEQUENCE	FROM N.A.		
RP	MEDLINE=93064711; PubMed=1437115;		
RX	Mitchell P.J., Cooper C.S.;		
RT	"The human tpr gene encodes a protein of 2094 amino acids that has		
RT	extensive coiled-coil regions and an acidic C-terminal domain.";		
RL	Oncogene 7:2329-2333(1992).		
RN	[2]		
REVIEWS	AND CHARACTERIZATION.		
RP	MEDLINE=95096166; PubMed=7798308;		
RX	Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,		
RA	Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;		
RT	"tpr, a large coiled coil protein whose amino terminus is involved in		
RT	activation of oncogenic kinases, is localized to the cytoplasmic		
RT	surface of the nuclear pore complex.";		

Db 935 YDEAREEVRLRQELTQL-HEDLKRAESDKNDSYKKRELERQRAKVIEVEKELEERYFD 993
QY 1256 D-----HGGIDLEWLRESPPDKADYLLSIR--GLGLKSVECVRLLLTLHNLAFFVDT 1305
Db 994 DSRLAECKRHGD---EMLR-----KIADLEKLRDGGNGGNGCTSSCEFER----- 1038
QY 1306 NVGRIAVRMGWVLPQIPESLQHLLELYPVLESIQKFL-WPRL---C--KLDQ----- 1353
Db 1039 --KRIAV-----LEVEVRKSMETIKSLEKFMEDRLQKDCADKLDREKERRM 1083
QY 1354 RTLYELHYQLITFGKVCTKSRPNCNACPMRGECRHFASAYASARLALPAPEE----- 1406
Db 1084 KAERDLEREI-----ARKNCGNCPERELESERSNVKREYOLDAEKVKFYKR 1133
QY 1407 -----RSLTSATIPVPPESFPVPAIPMIELPLPLEKSLASGAPSNRENCEPIIEE 1456
Db 1134 ELERDYLSSRYLTSSDP-DEKPLPNYTFPRIV-----EPLTTEDEPKPVE 1181
QY 1457 PASPGQECTEITESDIEDAYNEDPDEIPTIKLNIEQFGMTLREHMERNMELQEGDMSKA 1516
Db 1182 WPPSSDVTEPISSGV-----TPSVDAEPE-----HPQLSEYQTSVS 1218
QY 1517 LVALHPTTTSIPTPKLKNISRLRTEHQVYELPD SHR-----LLDGMKREPDPS 1567
Db 1219 QVAVTTP-----PKPETQISEYQDSELYSASNNTESKNVFSLAYLDDLKDL--DDIDE 1272
QY 1568 YLLAIWTPGET 1578
Db 1273 YLLNNINPEKT 1283

Search completed: September 27, 2002, 13:14:22
Job time: 1110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:54:47 ; Search time 122.1 Seconds
(without alignments)
2449.699 Million cell updates/sec

Title: US-09-840-743-2

Perfect score: 9089

Sequence: 1 MOSIMDSAVNATEQND.....PRPLMARLHPASKLNNKT 1729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:**
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	3949	43.4	1017	Q9L269	Q9L269 arabidopsis
2	2863	31.5	555	Q9L268	Q9L268 arabidopsis
3	1776.5	19.5	1309	Q9SR66	Q9SR66 arabidopsis
4	1559.5	17.2	1207	Q9SJQ6	Q9SJQ6 arabidopsis
5	1260	13.9	234	Q9L267	Q9L267 arabidopsis
6	999	11.0	274	Q9FTQ2	Q9FTQ2 oryza sativ
7	709	7.8	917	Q9498	Q9498 arabidopsis
8	458	5.0	119	Q94LX6	Q94LX6 euphorbia t
9	281	3.1	2151	Q9NG79	Q9NG79 trichomonas
10	278	3.1	5327	Q96891	Q96891 drosophila
11	253	2.8	17352	Q95YM2	Q95YM2 procambiarus
12	245	2.7	2254	Q9HCY0	Q9HCY0 homo sapien
13	241	2.7	2187	Q9H197	Q9H197 homo sapien
14	240.5	2.6	2938	Q61769	Q61769 mus musculu
15	239	2.6	2701	Q9Y520	Q9Y520 homo sapien
16	239	2.6	3574	Q9AUB4	Q9AUB4 arabidopsis

ALIGNMENTS

RESULT 1

Q9L269 ID Q9L269 PRELIMINARY; PRT; 1017 AA.
AC Q9L269;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL I12.1 KDA PROTEIN.
GN T32M21_160.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_taxid=3702;
RN SEQUENCE FROM N.A.
RP Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemck K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162875; CAB8562.1; -
KW Hypothetical protein.
SQ SEQUENCE 1017 AA; 4F7C313A891EC83F CRC64;

Query Match 43.4%; Score 3949; DB 10; Length 1017;
Best Local Similarity 99.9%; Pred. NO. 1.8e-246;
Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOSIMDSAVNATEQNDGSRQDVLEFDLNKTPQKPKRKFMPKVVVEGPKRKP 60

Db 235 MOSIMDSAVNATEQNDGSRQDVLEFDLNKTPQKPKRKFMPKVVVEGPKRKP 294

QY 61 RKPAELPKVVVEGPKRKPRAATQEKVKSETGSAKKNLKESATKKPANVDGMSNKP 120

|||||
295 RKPAPKVVVEGPKRPRKAKQEKYKSKETSAKKNLKESATKKPANVGDMNSKSP 354
Db
121 EVTLKSCRKALNFDLEPGDARQDSSEIIVQNSGGANSFSEIRDAIGGTNGSFLDSVSQ 180
Qy
355 EVTLKSCRKALNFDLEPGDARQDSSEIIVQNSGGANSFSEIRDAIGGTNGSFLDSVSQ 414
Db
181 IDKTNGLGAMNQPLEVSMGNQDPDKLSTGAKLARDQDPDLTLRNQCCQFPVATQNTQPFME 240
Qy
415 IDKTNGLGAMNQPLEVSMGNQDPDKLSTGAKLARDQDPDLTLRNQCCQFPVATQNTQPFME 474
Db
241 NOQAWLQMNKQNLIGFPNGNQPRMTIRNQPCCLAMGNOQPMYLGTPRPALVSGNQQLGG 300
Qy
475 NOQAWLQMNKQNLIGFPNGNQPRMTIRNQPCCLAMGNOQPMYLGTPRPALVSGNQQLGG 534
Db
301 POGNKRPLFLAHQTCPLPAGNOLYGSPTDMHQLVMSGTGQGHLLIKNOOPGSLIRGOQPC 360
Qy
535 POGNKRPLFLAHQTCPLPAGNOLYGSPTDMHQLVMSGTGQGHLLIKNOOPGSLIRGOQPC 594
Db
361 VPLIDQOQPATPKGFTHLNQMVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNGTTGTCQ 420
Qy
595 VPLIDQOQPATPKGFTHLNQMVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNGTTGTCQ 654
Db
421 RSRAPAYDSLODTHQGNKYILSHISNGCKKALPONSSSLPTPIAKLEEARGSRKQY 480
Qy
655 RSRAPAYDSLODTHQGNKYILSHISNGCKKALPONSSSLPTPIAKLEEARGSRKQY 714
Db
481 HRAMGOTEKHLNLAQQAQIAQSDQVERHNSSTCEVYLDAAKTKTKQVVOENLHGMPPPEVI 540
Qy
715 HRAMGOTEKHLNLAQQAQIAQSDQVERHNSSTCEVYLDAAKTKTKQVVOENLHGMPPPEVI 774
Db
541 ETEDDPTDGARKGNKNTASISGASKGNSPVKKTAKEKCIIVPKTPAKKGRAGRKKSVP 600
Qy
775 ETEDDPTDGARKGNKNTASISGASKGNSPVKKTAKEKCIIVPKTPAKKGRAGRKKSVP 834
Db
601 PAHASEIQLWQPTPKTPLSRKPKGKRKSTQDSGKARGPSGELLCCDSTAEIIRYMON 660
Qy
835 PAHASEIQLWQPTPKTPLSRKPKGKRKSTQDSGKARGPSGELLCCDSTAEIIRYMON 894
Db
661 LYLGGKEREQEQANVLYKGDCALVPYESSKKRPKPKYDIDDETRINWLLMGKDEKEG 720
Qy
895 LYLGGKEREQEQANVLYKGDCALVPYESSKKRPKPKYDIDDETRINWLLMGKDEKEG 954
Db
721 DEEDKKKEKWEERRVFRGRADSFIAHMLVQGD 756
Qy
955 DEEDKKKEKWEERRVFRGRADSFIAHMLVQGE 990
Db

RESULT 2

Q9LZ68 PRELIMINARY; PRT; 555 AA.

AC Q9LZ68;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 62.1 KDA PROTEIN.
GN T32M21.170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villaroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL162875; CAB85563.1; -.
DR InterPro; IPR003265; Endo_3c.
DR SMART; SM00478; ENDO3c; 1.
KW Hypothetical protein.
SQ SEQUENCE 555 AA; 62088 MW; A1B44BDDEL17FDC1E CRC64;

Query Match 31.5%; Score 2863; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 1.le-176;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 785 SSAPMSLAARPPPKLSSREDERNVRSVVVEDPGECLILNELIPEISWQEKVOHPDSMEVSG 844
Db 7 SSAPMSLAARPPPKLSSREDERNVRSVVVEDPGECLILNELIPEISWQEKVOHPDSMEVSG 66
Qy 845 VDSGSKQLRDCNSNGIERFNFLFKSIQNLEEEVSSQDSFDPALFQSCGRVGSQSCSKS 904
Db 67 VDSGSKQLRDCNSNGIERFNFLFKSIQNLEEEVSSQDSFDPALFQSCGRVGSQSCSKS 126
Qy 905 DAEPPTTRCETKTVSGTSQSQSVOTGSPNLSDEICIQGNERPHLYEGSGDVQKQETTNAQK 964
Db 127 DAEPPTTRCETKTVSGTSQSQSVOTGSPNLSDEICIQGNERPHLYEGSGDVQKQETTNAQK 186
Qy 965 KPDLEKTMNWKDSVCFQGPNDTNWQTPSSSYEQCATRQPHVLDIEDFGMQGGLGYSW 1024
Db 187 KPDLEKTMNWKDSVCFQGPNDTNWQTPSSSYEQCATRQPHVLDIEDFGMQGGLGYSW 246
Qy 1025 MSISPRVDRVKNKNVPRFFRQGGSVPREFTGQIIPSTPHELPGMLGSSSAVQEHQDD 1084
Db 247 MSISPRVDRVKNKNVPRFFRQGGSVPREFTGQIIPSTPHELPGMLGSSSAVQEHQDD 306
Qy 1085 TOHNOQDEMKNKASHLQKTFDLNLSSECLTRQSTKQNTDGLPRDRTAEDVVDPLSN 1144
Db 307 TOHNOQDEMKNKASHLQKTFDLNLSSECLTRQSTKQNTDGLPRDRTAEDVVDPLSN 366
Qy 1145 NSSLQNTILVESNSSNKEQTAVEYKETNATILREMKGTLDGKKPTSQWDSLRLKDVQEGNEG 1204
Db 367 NSSLQNTILVESNSSNKEQTAVEYKETNATILREMKGTLDGKKPTSQWDSLRLKDVQEGNEG 426
Qy 1205 ROERNKNMDSIDYEAIRRASISEISEAIKERGMNMLAVRIKDFLERIVKHGGLDLEW 1264
Db 427 ROERNKNMDSIDYEAIRRASISEISEAIKERGMNMLAVRIKDFLERIVKHGGLDLEW 486
Qy 1265 LRESPDPKADYLLSIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGVLPLOLPE 1324
Db 487 LRESPDPKADYLLSIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGVLPLOLPE 546
Qy 1325 SLQLHLEL 1333
Db 547 SLQLHLEL 555

RESULT 3

Q9SR66 PRELIMINARY; PRT; 1309 AA.
AC Q9SR66;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T22K18.18 PROTEIN.
GN T22K18.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";

RL Submitted (JAN-2001) to the ENBL/GenBank/DBJ databases.
 DR EMBL; AC010927; AAF04422.1; -.
 DR InterPro; IPR003265; Endo.3c.
 DR InterPro; IPR003651; FeS_bind.
 DR Pfam; PF00730; HHH-GPD; 1.
 DR SMART; SM00478; HHH-GPD; 1.
 DR SMART; SM00525; FeS; 1.
 DR SMART; SM00525; FeS; 1.
 SQ SEQUENCE 1309 AA; 148451 MW; 3B21E4439BAA673 CRC64;

Query Match	19.5%	Score 1776.5;	DB 10;	Length 1309;
Best Local Similarity	36.0%;	Pred. No. 6.3e-106;		
Matches 451;	Conservative 154;	Mismatches 300;	Indels 347;	Gaps 41;

QY	536	PPEVIEIDD-PTDGARKKNTASISKGASKG-----NSSPVKTAEBKECIVPKTPA	587
Db	332	PGFMESEEDRPD-----SOISLQGRSITMATPRNFRSLAKLLQR---II---PS	377
QY	588	KKGRAGRK---KSVPPPAHASEIOLWQTPPKTPLRSKPKGKRKSIODSG-----KA	638
Db	378	KRDKGCKLPRGLUKLTVASKLQ-----KVFRKRSQRNRVASQFARILDLQWR	429
QY	639	RGPSGELLCO-----DSIAEIIYRMONLYLGDKRE-----QEQNAMVLYKGDGAL	684
Db	430	QNPTGTSLADIWERSLTIDAITKLFEE-----DINKEGCLPHNRETAIILYKKS---	480
QY	685	VPYESK-----RKPRPKVDIDETTRIWNLLMGKDEKEDEKDKKKKEWEEERRV	738
Db	481	--YEEOKAIVKYSKKOKVOLQDPETSRVKMLLMSSID--CDGVGDSDEKRWKEERNM	537
QY	739	FRGRADSFIAIRMLHVOGDRFRFPWKSGSVDSVYGVELTQNVSDHLSASFMSLAARFP	798
Db	538	FHGRANSFIARMRVQGNRTFPSWKSGSVDSVVGVELTQNVADHSSSAYMDLAAEFPE	597
QY	799	LS-----SSREDERNRVSVEDDEGCILNLNETPSWQEKVQHPSPDMEVSGVDSGSKEQLR	854
Db	598	WNFNKSGCHEE-----WGSV-----	613
QY	855	DCSNSGIERFNFLKESIQNLLEEVLSSQSDFPDPAIFQSGRGVSCSCSKSDAEFTTICE	914
Db	614	-----TQETILNL-----DPRTGVSTPRIN-----PTRVI	639
QY	915	TKTVGTSQSVQFGSPNLSDEICLQGNRPHLYEGSGDVOKOETTNVAQKKFPLDKTMW	974
Db	640	IEEIDDDENDI-----DAVCSQ-----ESSKTSDSITSADQSKTM-----	675
QY	975	KDSVCFQGNPDNTNWOTTPSSSYEQACATRQPHVLDIEDFGMGEGGLYSWMSISPRVDV	1034
Db	676	-----LLD-----PNTVL	684
QY	1035	KNKNVPRFRFGSGVPREFTGIIIPSTPHELPGMLSGSSSAVOBHOJDDTOHNOQDEM	1099
Db	685	MNEQVDSQWYKKGCHIP--YTDDL-----NDL--SQGISWVSSA--STHCELNLNEVPPEVE	735
QY	1095	KASHLQXTFDLLNLSSECLTROSSSTKQNTDGCCLPRDRTAEDVDVPLSNSSQLONILVE	1154
Db	736	LCSHOODPESTIQTODQOESTRTEDVKN-----	764
QY	1155	SNSSNKEQTAVEYKETNATILREMGKTLADGKKPTSQWDSLKDYEGNEGROERNKNND	1214
Db	765	-----RKPTTSKPKKSKESAKSTQKSDWDSLDRKEAESGGRKERTERTMD	813
QY	1215	SIDYEAIRRASISEISAEIKERGMNNMLAVRIKDFLERIVKDHGGIDLEWLRESPPDKAK	1274
Db	814	TVDWDLARCTDVUKIANIIKRGNNMNLAEIKAFNLRLVKKHGSLDLBWLKDVPPDKAK	873
QY	1275	DYLLSIRGLGLKVECVRLTLHLNAFFVDNTNNGRIAVRMGWVPLOPLPESLQLHLLEY	1334
Db	874	EYLLSINGLGLKVECVRLSLSLHQAFFVDNTNNGRIAVRLGWVLPLOPLDQLMHLE---	931
QY	1335	PVLESIOKFLWPRCLKDQRTLVELHYQLITFGKVFCTKSRRPNCNACPMRGECRHFASAY	1394
Db	932	-----LYELVHMTFGKVFCTKVKVKNACPMKACRYSSAR	970

Qy	1395	ASARLALPAPERLSATSIPVPPSPFPVAI---PMIELPPLPEKSLASGAPSNRENCE	1451
Db	971	ASARLALPEEESDRTSMITHERRSKRPVVNFRPSLFLYQKEQE-----AQRSQNC	1025
Qy	1452	PIIEEPASFOCETETIESDIED-----AYNEDP---DEPTIKLNIEQFGMTLRE	1500
Db	1026	PIIEEPASPE---PEYIEHDIEDYPRDKNNVGTSEDPWENKDVPTIILNKE-----	1074
Qy	1501	HMERNMEL---QEGDMSKALVALHPTTTSIPTPKLNISRLETHOVYELPDSHRLDGM	1557
		: : : : : : : : : :	
Db	1075	-AGTSHDLVVNNKAGTSHDLVLSLYAAAPRRKKLUIKEKLRTEHHPELDPHSHLEGF	1133
Qy	1558	DKREPDDPSYLLAIWTPGETANSAPPEQKCG-CKASGKMCFCDBTSECNSLEANSQT	1616
		: : : : : : : : : : :	
Db	1134	ERREAEDIVYLLAIWTPGETVNSQPPKQRCALFESNNTLCNENKCFQCNKTRRESQT	1193
Qy	1617	VRGTLLIPCRATMRGSPFLNGTYFYQNELFADHESLAKPIDVPRDWIWDLPRTTYFGTS	1676
Db	1194	VRGTILIPCRATMRGSPFLNGTYFYQNEVFADHDSSINPIDVPTELIWDLARRVAILGSS	1253
Qy	1677	VTSIFRGLSTEQTOFCFWKG-FVCVGRGFPQKTRAPRPMARLHPSPASKLKN	1727
		: : : : : : : : : :	
Db	1254	VSSICKGLSVEAIKYNFQEGMLICL---TLKT---PLVSNIEF---MKRTNN	1296
RESULT	4		
Q9SJQ6			
ID	O9SJQ6	PRELIMINARY:	PRT: 1207 AA.

RESULT	4
Q9SJ06	
ID	Q9SJ06 PRELIMINARY; PRT; 1207 AA.
AC	Q9SJ06;
DC	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DI	01-DIC-2001 (TrEMBLrel. 19, Last annotation update)
DE	AT2G36490 PROTEIN.
GN	AT2G36490.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid s II; Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_TaxId=3702;
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=CV. COLUMBIA;
RC	MEDLINE=20083487; PubMed=10617197;
RX	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA	Salzberg S.L., Fraser C.M., Venter J.C.;
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT	thaliana.";
RL	Nature 402:761-768(1999).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA;
RC	Lin X.;
RL	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RL	EMBL; ACC06919; AAD24633.1; -.
DR	Interpro: IPR003651; Fes_bind.
DR	SMART; SN00525; FES; 1.
SQ	SEQUENCE 1207 AA; 135916 MW; ADD86C93EE94DDF67 CRC64

	Query Match	17.2%;	Score	1559.5;	DB	10;	Length	1207;
	Best Local Similarity	28.5%;	Pred. No.	5.9e-92;				
	Matches	480;	Conservative	199;	Mismatches	366;	Indels	641; Gaps
								61;
Qy	13	TEATEQNDSQRDVLFEFLNKTPOQSKRRKFMFKPVVVVEGPKRKRPAELPKVYVE	72					
Db	65	TEEVESLSVSNNVAE--OILKTP-EKPRKKHR--PKVRREAKPREEKRPAPRKSVYTD	120					

```
Qy 73 GKPKRKAATQEKVK-SKETGSAKKKLNKESATKPKANVGMNSKSPVETLKSCKRAL 131
Db 121 QESKTPKRYVRKKVEYKSD-----QDAPVSSAAVETSRPK-----RLCRVL 167
Qy 132 NFDLEN-----PGDAQ-GDSESEIVQNSGANSFSELRDAIGTNGSFSLDSVSDIKTN 185
Db 168 DFEAENGNGNDGIDREAGEMESALQEKQLDSGN-QELKDCLE-----LSAPSTPKRRK 219
Qy 186 GIGAMNOPLEYSMGNOPDK-----LSTGAKLARDQOQDPLLTRN-----QOQOFP 229
Db 220 SQGRK-----GVQPKKNSNLEEVDISMAQAARQOQPTCCDMNLSGIYDEQCDY- 271
Qy 230 VATONTOPMENQAWLQMKNLQIGFPFGNOQPRMTIRNOQPCIAMGNQOQMYLIGTRP 289
Db 272 -----QKHWLYSPNL-----QOQGM--RYDAIC----- 293
Qy 290 ALVSGNOQLGPGQGNKRPFLNHOTCLPAGNQLYGSPTDMHQLVMSTGGQGHLLKNNQ 349
Db 294 -----SKVFSQQHNYVSF--HATCYSSSTLSQSANRY-----LTVEERR 331
Qy 350 PGSILRGQOQCVPLIDQOATP-----KGFTHLNQMWATSMSSPGLRPHSQSV 398
Db 332 EGIFQGRQESLNLVDKIDPIKKTTHGAFRPNLSMKNLV-----EV 376
Qy 399 PTTYLHVSVSRILNGTTGTQCRAPAYDSLOQDIHOGNKYILSHEISNGNGCKKALPQ 458
Db 377 PE--HL-----TSGYCKPQ-----QNNKILVDTRVT----- 401
Qy 459 NSSLPTPIWAKLEARGSKROYHRAMQOTEKHDLNLAQIAQSODVERHNSSTCEVLYDA 518
Db 402 -----VSKKPKTKSEKQ----- 414
Qy 519 AKTKIQKRVQVQENLHGMPEVIEIEDDPTDGARKGKNTASISKGASKGNSPVKTAKE 578
Db 415 --TK-QKNLLPNLCRFPFSTGLSPD----- 437
Qy 579 KCIVPKTPAKGRAGRKSVPPPAHASIQWLQPTPKTPLSRKPKGKGRKSQDSOKA 638
Db 438 -----ELW-----KRNNSI----- 446
Qy 639 RCPGSELLCQDSIARLIYRMQNLVILGDKEREQONAMVLYK-----GDGALVPE 688
Db 447 -----ETISEL-----LRLLDINREHSEFALVPTMNSQIVLFGGAGAIAPVT 490
Qy 689 S-KRKRPRPKVIDDETTRINWLLMGKDEKDEKDEKKEKWEERVRFRGRADFSI 747
Db 491 PVKPRPRPKVLDDETRVWKLKLL-ENINSEGVDSDEQRAKWEERVRFRGRADFSI 549
Qy 748 ARMHLVQDGRFRSPKWSGVSDVIGVFLTONVSDHLSSAFMSLAARPP-PKLSRSSREDE 806
Db 550 ARMHLVQDGRFRTPKWSGVSDVVGVLTONVSDHLSSAFMSLASQFPVPFVPSNFD- 608
Qy 807 RNRSVVVDEPGCLILNLEIPSMQ-----EKVQHPSDMEVSGVSDSGSKQLRDCSNS 859
Db 609 -----AGTSSMPSTQITYLDSEETMSSPPDHNSSVT----- 640
Qy 860 GIERENFLEKSIQNLDEEVLSQSDSDPAIFQSGRVCSCSKSDAEFTTRCT--KT 917
Db 641 -----LKNTQPDKEKYVPSNET-----SRSSSIAISAHESVDKT 676
Qy 918 VSGTS--QSVQTSQPNLSDEICLOGNERPHLYEGSGDVOK-OETTNVAQKPDPLEKTNW 974
Db 677 TDSKEYVSDRKS-----SVEVDKTDKCRVNLNLPSPEDSALTC 716
Qy 975 KDSVCFQGPNDNTNNQWTFSSSYEQCATRQPHVLIDIEFGMQGELGVSWMSISPRVDV 1034
Db 717 QHSMVSDAPQNT---ERAGSSS-----EIDLEGE-YRTSFMKL----- 750
Qy 1035 KNKNVPRFRFGGVSVPREFTGQIIPSTPHLPGLMGLSGSSSAVOEHQDDTOHNOODEMN 1094
Db 751 -----LOGVOVSLSDNSQVSPN-----MSPGDCSSEIKGFQ-----SMKEPT 787
Qy 1095 KASHLQKTFDLNLSSEE-CLTRQSTKQNTDGLCLPRDRTAEDVVDPLSNSSNLQNLV 1153
```

```
Db 788 KSS-----VDSSEPGCCSQ-----DG-----DVL----- 807
Qy 1154 ESNSNNKEQTAVSEYKETNATILREMKGTLADGKKPTQOWDSLKRDVEGNEGRQERNKNN 1213
Db 808 ----SCQKPT--LKEGKKVLE-----EKKAFDWDCLREAAQARAGIREKTRSTM 852
Qy 1214 DSDIYEAIRASTSEISEAIKERGMNNMLAVRI-----KDFLERIVKDHGGIDLE 1263
Db 853 DTVDWKAIRAADKVEAETIKSRGMNHKLAERIQYLTNNKIMQGFQDLRLVNDHGSIDLE 912
Qy 1264 WLRESPPDKADYLLSTRGLKSVCEVRLTLTLNLAFFVDTNVTAVRMGWVYPLQPLP 1323
Db 913 WLRDVPPDKAK----- 923
Qy 1324 ESLQLHLELYPVLESIOKFLWPRCLKDQRTLYELHYQLITFGKVFCTKSRPNCNACPM 1383
Db 924 -----YELHYQMTITFGKVFCTKSKPNCNACPM 950
Qy 1384 RGECHRFASAYASARLALPAPERSLTSATIPVPSPFPVPAIMIEPLPLEKSLASGA 1443
Db 951 KGECHRFASAFASARLALPSTEKGMGTDPKNPLP-----LHLPEPFQREQGSEV 999
Qy 1444 PSNREN-----CEPIIEEPASPGQECTEITESDIEDAYYNEDEPDEPTIKLNIQEQGMT 1497
Db 1000 VQSEPAKVTCCPEPIIEEPASPEPETAESVIADIEEAF-EDPEEPTIKLNMDAFTSN 1058
Qy 1498 LREHWRNEMELQEGDMSKALVALHPTTTSIPTPKLNISRLRTEHQVYELPDSHRLLDGM 1557
Db 1059 LKTIMEHKELODGNMSSALVALTAETASLPMPLKLNISQLRTEHRYVELPDEHPLLAQL 1118
Qy 1558 DKRPDDPSPYLLAIWTPGETANSQAQPEQKCGGKASGKMGCFDETCSECNSLRANSQTV 1617
Db 1119 EKRPDDPCSYPYLLAIWTPGETADSTQPSVSTCIFAQNGMLCDEETCFSCNSIKETRQIV 1178
Qy 1618 RGTLLI 1623
Db 1179 RGTILV 1184
```

RESULT 5

```
Q9L267 PRELIMINARY; PRT; 234 AA.
AC Q9L267;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 26.6 KDA PROTEIN.
GN T32M21_180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysschaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villaroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lencke K., Mayer K.F.X.
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162875; CAB85564.1;
KW Hypothetical protein.
SQ SEQUENCE 234 AA; 26649 MW; B395392E066EB297 CRC64;
```

Query Match 13, 9%; Score 1260; DB 10; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.2e-73;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1496 MTLREHMERNELOEGDMSKALVALHPTTTSIPTPKLNISRLRTEHQVYELPDSHRLLD 1555
DB 1 MTLREHMERNELOEGDMSKALVALHPTTTSIPTPKLNISRLRTEHQVYELPDSHRLLD 60
QY 1556 GMDKREPDPSPYLLAIWTPGETANSQPPQKCGKASGKMCDFETCSECNLSREANSQ 1615
DB 61 GMDKREPDPSPYLLAIWTPGETANSQPPQKCGKASGKMCDFETCSECNLSREANSQ 120
QY 1616 TVRGTLIPCRBTAMRGSPFLNGTYFQVNELFADHESLKPIDVPRDWIWLPRRTVYFGT 1675
DB 121 TVRGTLIPCRBTAMRGSPFLNGTYFQVNELFADHESLKPIDVPRDWIWLPRRTVYFGT 180
QY 1676 SVTSIFRGLSTEQFOCFWKGFVCVRGFEQKTRAPRPLMARLHPFASKLNKNT 1729
DB 181 SVTSIFRGLSTEQFOCFWKGFVCVRGFEQKTRAPRPLMARLHPFASKLNKNT 234

RESULT 6
Q9FTQ2 PRELIMINARY; PRT; 274 AA.
AC Q9FTQ2
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0665D10.14 PROTEIN.
GN P0665D10.14
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0665D10."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002861; BAB16489.1; -.
SQ SEQUENCE 274 AA; 30951 MW; D7C826BD3982592D CRC64;

Query Match 11.0%; Score 999; DB 10; Length 274;
Best Local Similarity 68.6%; Pred. No. 1.1e-56;
Matches 179; Conservative 38; Mismatches 42; Indels 2; Gaps 2;

QY 1469 ESDIEDAYNEDPDEIPTIKLNIEQFGWTLREHME-RNNELOEGDMSKALVALHPTTTSI 1527
DB 3 ECATEDSFV-DDPEEPTIKLNFEFTQNLKSYMOANNIEEDADMSKALVAITPEVASI 61
QY 1528 PTPKLNISRLRTEHQVYELPDSHRLLDGMDKREPDPSPYLLAIWTPGETANSQAPPEQ 1587
DB 62 PTPKLNISRLRTEHQVYELPDSHRLLDGMDKREPDPSPYLLAIWTPGETANSQAPPEQ 121
QY 1588 KCGKASGKMCDFETCSECNLSREANSQTVRGTLIPCRBTAMRGSPFLNGTYFQVNELEFA 1647
DB 122 VCNSENGELCASCTCFSCNSIREAQAKVRGTLIPCRBTAMRGSPFLNGTYFQVNEVEFA 181
QY 1648 DHESLKPIDVPRDWIWLPRRTVYFGTSVTSIFRGLSTEQFOCFWKGFVCVRGFEQK 1707
DB 182 DHDSRRNPIDVPRSWIWNLPRTVYFGTSVTSIFRGLSTEQFOCFWKGFVCVRGFEQK 241
QY 1708 RAPRPLMARLHPFASKLNKNT 1728
DB 242 RAPRPLMARLHPFASKLNKNT 262

RESULT 7
O49498 PRELIMINARY; PRT; 917 AA.
ID O49498
AC O49498
DT 01-JUN-1998 (TrEMBLrel. 06, Created)

```

```

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 106.3 KDA PROTEIN.
GN F28A23.180 OR AT4G34060.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
RA Herzl A., Neumann S., Hoheisel J., Jesse T., Heijnen L., Vos P.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021961; CAAL17566.1; -.
DR EMBL; AL161584; CAB80123.1; -.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; Fes_bind.
DR Pfam; PF00730; Hhh-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; FES; 1.
KW Hypothetical protein.
SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;

Query Match 7.8%; Score 709; DB 10; Length 917;
Best Local Similarity 21.4%; Pred. No. 3.5e-37;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

QY 651 IAEIIRYRMQNLVGLGDKEREQEQONAWLVKGGALVPVYESKKRPRKVDIDDETRLNWL 710
DB 388 IAKIKDMGRLLNKK-----VTTMIKADKKLV-----TAKNLDLPETIKEMDV 431
QY 711 LMKGDEKEGDEEDKKKKKWEERVFRRGRADSFARMHLVQGRFRFPWKGSVVDSV 770
DB 432 LM--VNDSPRSYDDKTEAKKKEREIFQTRIDLFNRMHRLQGNKKFKQWKGSVVDSV 489
QY 771 IGVFLTONVDHLSSSAFMSLAARFPKPLSSSRDEDRNRSVVVEDPEGCILNLNEIPSW 830
DB 490 VGVFLTONVTDYLLSSNAFMSVAARFP-----VDAREGLSYITEEPQ----- 530
QY 831 QEKVQHPSDMEVSGVDSGSKQLRDCNSGLIERNFLEKSIQNLEEVLSQDSFDPALF 890
DB 531 -----DAKSSCCI----- 538
QY 891 QSGGRVSGSCSKSDAEFTTRCETKTVSGTSQVQTSQVQTSQVQTSQVQTSQVQTSQVQTS 950
DB 539 -----ILSDE----- 543
QY 951 GDVQKQETTVAOKKPPLEKTMNWKDVCFGQPRNDTNWQTPSSSYEQCATRPHVLDI 1010
DB 544 ----- 543
QY 1011 EDFGMQGEGLGYSMWSISPRVDRVKNVPRFRFQSGSVPREFTGQIIPSTPHELPMGM 1070
DB 544 ----- 543
QY 1071 LSGSSSAVQEHQDDTQHNQODENMKASHLQKTLFDLLNSSECLTROSSTKQNTDGLCP 1130
DB 544 ---SISKVEDHEN-----TAKRKEKGTGII-- 565
QY 1131 RDRTAEDVDPPLSNSSQLNILVESNSSNKEQTAVEYKETNATILREMKGTLDAGKKPTS 1190
DB 1131 ----- 1190

```

```
Db 566 -----EDEIVD----- 571
Qy 1191 QWDSLRKDVEGNEGRQERNKNMDSIDYEAIRRASISEAIEKRGMMNNLAVRIKDFL 1250
Db 572 -WNNLR--MYTKEG--SRPEMHMSVNVSDVRLSQNVLETIKRGQFRILS----- 620
Qy 1251 ERIVKDHGGIDLEWLRSPDPKADKDYLLSIRGLGLKSVCEVRLTLHLNLAFFVDVTNNGRI 1310
Db 621 ERIL-----VDTNVGR 632
Qy 1311 AVRGMVPLQPLPSLQHLLELPVLESIOQLFWPLRCLCKLDQRTLYELHYQLITFGKVF 1370
Db 633 AVRGLVPLEPLPAGVQMH-----QLFEYELHYQMITFGKVF 669
Qy 1371 CTKSPNCPNACPMRGECRHFASAVASARLALPAPAEERSLTSATIPVPPESP-----PPVA 1425
Db 670 CTKIPNCPNACPMKSECKYFASAVSVSKVLLESPEEK-----HEPNTFMAHSQDVA 722
Qy 1426 IPMIELPLPLEKLSAGSPNRENCEPIIEEPASPGOECTEITESDIEDAY-----NEDPD 1482
Db 723 VDMTSINLVECVSSGCSQDAICYKPLVFPSPSPRAIPEST--DIEDVPFMYQSYA 780
Qy 1483 EIPTKLNIEQFGWTLREHM--ERNMELQEGDMSKALVALHPTTISITP---KLKNISR 1537
Db 781 SVPKIDFDLALKKSVEDALVISGRMSSDEEISKALVIPTPENACIPKPPRKMYYNR 840
Qy 1538 LRTEHOVYELPDHRLDGMKREPDPSPYLLAIWTPGETANSAPPEQKCGGKASKM 1597
Db 841 LRTEHVYVLPDNEHLLHDFERKLDPSPYLLAIWQP----- 878
Qy 1598 CFDETCSECNLSREANSQTVRGTLILPCRTAMRGSFPLNGTYFQVNFELFADHESLAPID 1657
Db 879 ----- 878
Qy 1658 VPRDWIDLPRRTVYFGTSVTSIPRGLSTEQIQCFWKGFVVCVRGFEQKTRAPRPLMARL 1717
Db 879 -----GFLCLRAFDKQDRKPELVRL 900
Qy 1718 HFP 1720
Db 901 HTP 903

RESULT 8
Q94LX6 PRELIMINARY; PRT; 119 AA.
AC Q94LX6.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE FPPSYNTHASE1 (FRAGMENT).
GN ETPPSSI.
OS Euphorbia tirucalli.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.
OX NCBI_TaxID=142860;
RN [1].
RP SEQUENCE FROM N.A.
RA Asada T., Fukusaki E., Kobayashi A., Sando T.;
RT "Prenyltransferase of Euphorbia tirucalli.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051105; BAB47587.1; -.
FT NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13771 MW; FBDCE4CD0C5B01 CRC64;

Query Match 5.0%; Score 458; DB 10; Length 119;
Best Local Similarity 78.7%; Pred. No. 2.8e-22;
Matches 85; Conservative 10; Indels 0; Gaps 0;

Qy 1214 DSIDYEAIRRASISEAIEKRGMMNNLAVRIKDFLERIVKDHGGIDLEWLRSPDPKA 1273
```

```
Db 12 DYLDWEAVRCADVDEIANTIKERGMMNVLAKRIRKIDFNRLVREHGNIDLEWLRDPDPKA 71
Qy 1274 KDXLLSIRGLGLKSVCEVRLTLHLNLAFFVDVTNNGRIAVRMGWVPLOP 1321
Db 72 KEYLLSIRGLGLKSVCEVRLTLHLNLAFFVDVTNNGRIAVRMGWVPLOP 119

RESULT 9
Q9NG79 PRELIMINARY; PRT; 2151 AA.
AC Q9NG79.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL 232.4 KDA PROTEIN.
OS Trichomonas vaginalis.
OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonas.
OX NCBI_TaxID=5722;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=#202;
RA Delgaty K.L.;
RT "The Molecular Identification and Characterisation of a Candidate Gene
for Cell Detaching Factor from Trichomonas vaginalis.";
RL Thesis (2000), University of Ottawa, Ottawa, ON, Canada, In press.
DR EMBL; AF257323; AAP67197.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2151 AA; 232366 MW; F2EF91C5D9C60C82 CRC64;

Query Match 3.1%; Score 281; DB 5; Length 2151;
Best Local Similarity 19.7%; Pred. No. 5.6e-09;
Matches 367; Conservative 256; Mismatches 705; Indels 536; Gaps 88;

Qy 10 VNATEATEQNDGSRQDVLEFDLNTPOOKPSRKRKPMKVVEG-----KPKRK 59
Db 372 LNLPKSPSSQTSSEKPKLSNLGKSPSNQSTEEK---PKQLGGIKLGQTSQNKSEK 427
Qy 60 PRKPAELPKVVVEGPK-----RKPKAATQEVKSKETGSAKKKNLKESATKKP---- 109
Db 428 PKLSNLGKSPSTEEKPKLSNLGKSPSNQSTEEK---LGFNLPKAPSNQTEKPKLGT 484
Qy 110 ----ANVGD--MSNKSPEVTILKCRKALNFLENPGDARGDSEITVQNSGANSFSEI 163
Db 485 GGISLNLGNKPQSEKPKLSLGGIKLA-----QSPSNSNEEKPKLSNLKSPSNQSTEE 539
Qy 164 RDAITGNTGFLDSVSDIDTNGLGAMNQPLEVSMGNOPD-----KLSTGA-KIARDOQP 217
Db 540 KPKLGFNLPKAPSNQTEKPKLGTGG---ISLNLGNKPQSEKPKLSLGGIKLA--QSP 593
Qy 218 DLLTRNQCOPPVATQNTQPFMENOQAWLQMKNOLIGFP--FGNQOPRMTIRNOOCLAM 275
Db 594 S-----NSNEEKPKLSNLKSPSNQSTEEKPKLQGLGKLNGLNKPQTEQTTEKPKLQ 649
Qy 276 G-----NQPMYLIPTRPALVSGNQQLGG-----PQGNKRPIFLNHTCLPAGNQLY 323
Db 650 GGIKLGQSPSNST--EEKPKL-----QLGGIKLNLGSKPQTEKPKL-----QLG 692
Qy 324 GSPFDMHQLVMSTGG-----QQHGLLNKQPGSLIRGOQPCVPLIDQ 366
Db 693 G-----IKLGTGGISLNLGNKPQSEKPKLQGLGKLGNSQP-----NQ 731
Qy 367 QPATPKGFTLHNQVATSMSSPGLRPHSQVPTTYLHVESVRIANGTTGTCCORSRAPA 426
Db 732 PLEKPKSGINLN-----LGKSQP-----SSEKPKLGLN-----GKSPSNTEKPK 774
Qy 427 YDSLQQDIHQGNKYILSHEISNGCKKALPQNSSLPTIPMAKLEEARGSKROYHRAMGQ 486
Db 775 LGTGGISLNLGNK-----PQTEKP----- 794
Qy 487 TEKHDLNLAQOIA-QSQDVERHNSSTCVYLDAAKTKIQ-KVVOENLHGMPPVEIED 544
```

Db 795 --KLSNLPKSPSNQONQSTE-----EKPQLQLGGLKLNIGNKPQTETQTEE 838
Qy 545 DP---TDGARKGNKTASISK-----GASKGNSSPVAKTAKEKCIVPKTPAKKGRAGR 594
Db 839 KPKQLGGLKLNLSKSTQTEKPKFQLGGIKLQGPSNSTEKPKLSL-----GG 888
Qy 595 KSVPPPPAHASEIQWQTPPTPKTPLSR--SKPK-----GKRKS-----IQDSG 636
Db 889 IKLAQSPNEEKPKL--SPNLKPSNQSTQTEKPKLQLGGLKLNLSKSTQTEKPKLSEGG 947
Qy 637 KARGPSGELLCCODSTAETIIRYQNLGLDKEREQONAM-----VLYKGDALVPY 687
Db 948 IKLGNVSSQTSDEKPKLGLGGLISFNQKQTEKPKLSSSQNVKPTLGLGGITLQOQ 1007
Qy 688 ESKKRKR-----PKVDIDDETRITWNLGKG-----DEKEGDEKDK----- 726
Db 1008 TSEENKPKLGNLPKSNQTEKPK-----LGTGGISLNLGNKPKQSEKPKLQLGGLKLAH 1062
Qy 727 KKEKWEEERRVFRGRADSFTARMHVGDRFRSPWKGSVVDS--VIGV--FLTQNVSDHLS 784
Db 1063 SPSNQSTEEK-----PKLSNLGK---SPSNQSTEEKPKLGLGGLKLNLSGHLQ 1108
Qy 785 SSAPMSLAARFPKLSRSSREDERNVRSVVVDEPECIILNLAIEPSWO--EKVOHPSDMEVS 843
Db 1109 SDE-----KPFSLGMLKGVPSNESQEPKFTLNIKPLPSNQAEQNDGPKLGLG 1160
Qy 844 GVD--SGSKQLRDCNSGI-----ERFNFLEK-----SIQNLDEEVLS----- 881
Db 1161 GISLNLQSKKEENEKPKLGIAPKQIQSNOQNKIEPKKPVNTSVTKTEKASAPGVKLDL 1220
Qy 882 -----QDSFDPALFOSCGRVGSCSKSDAEPPTTRCETKTVGVSQSVQTSQP 930
Db 1221 KLPKKGFLMRQPSKSLSEVSNDSQNLSSFQGIHVTLPERKEENQPLSFQO--IHVTL 1279
Qy 931 NLSDEICLQGNRPHLYEGSGDV-----OKOETTNVAOKKPDLEKTMWKKDSVCFQO--- 982
Db 1280 EKT-----QKRENOPLSSFQGIHVTLPOKQONEKVTQKVTDEKSL-----SFGQIHV 1329
Qy 983 --PRNDTNQ--TTSSSYEQCATRPHVLDIEDFGMQE-----GLGYSWMSISPRVDRYKN 1036
Db 1330 TIPEKLANQOEKPLSTPEQI-----HVSIEKTKQOEKISSFQGIHVSIPEKLOKQEE 1384
Qy 1037 K-----NVPREFFRQGG-----SVPREFTGQIIPSTHEL-----PG 1068
Db 1385 KTSFEGQIHVTIPEKLOKQEGSVIDKLP--EKTQELIKNMPESEINVKIPEQVHQSQSNPL 1443
Qy 1069 MGLSGSSSAVOEHODDTQHN-----QODEMKNASHLQKTFDLN-----SSECLT 1115
Db 1444 VPLGVSLNDVAKSQNTQNKPDLSLKPQKSEENKPS-----LGLKLPKPNFSLKPPQ 1499
Qy 1116 ROSSTKQNTDGLPRDRTAEDVDPLSNSSLO--NILVESNSNKEQTAVEYKRETNATI 1174
Db 1500 SQTESKPNLSSLLPPKQTL--GLKLPTSSNGLKLPKLPQNRRNQNEAKEVKET---- 1554
Qy 1175 LREMGTLADGKKPTSQWDSLRKDVEGNEGROE-----RNKNMDSIDYEATRASSIS 1230
Db 1555 -----KEYKAKEPKKELGFKSDFKIEGASPKGINLSSGKF-- 1591
Qy 1231 EAIKRGHNNLAVRIKDFLERIVKDHGGIDLEMLRSPDPKA--KDYLL-----SI 1280
Db 1592 EYPSGKSFDDMIKQVKAENKPKFSNLSNLSNLTPLSNAPKPIALSISNCGKKPSL 1651
Qy 1281 RGLGLSKVECVRLTLNLHIAFPVDTNVGRIAVRMGWVPLQPLPESLQLHLLLEXPVLES 1340
Db 1652 AGLNISKSSLASVINALDKSGD--DVPKALSIDKLKKEPPPP-----PTPAV 1700
Qy 1341 OKFLWPRLC-----KLDQRTLYELHQLITFGVCTKSRPNCNAC--PMRGEORHFAAYAS 1396
Db 1701 TK---PALATEKEIRESSIADV--TPLAMITKINSK--KPDFSALKPNLGNKQSSSSNEQ 1755
Qy 1397 ARLALPAPEERSLTSATIPVPPEFPFPAIPMIELPLPLEKSLASG-----APSNRENC 1450
Db 1756 KPLANKPNFSALKLPQKSEETKPNLSSLLPPKQTLGLKLPKLTSSNSGKLPKLPKPNQNOQ 1815

Qy 1451 EPIIEE-----PASPQOECTEITESDIEDAYYN-----EDPD----- 1482
Db 1816 NOEVKESKEVRAKEEPKELGFKSDFKIEGASFKGINLSSSKFEPVPSGKSFDDMIKOKA 1875
Qy 1483 --EPTTIKNTIEQGMTLREHMERMELO-----EGMSKALVALHPTTTSIPTP 1530
Db 1876 KTEOPKTEBNEPKFSFNLNLNLTPLSNAPKPIALSLSNDSNKGQVQOGALTIKPIP-- 1933
Qy 1531 KLNKISRLRTEHQYVELPDSHRLLDGMDKREPDDOPSPVLLAIWTPTGETANSQOPEQKCG 1590
Db 1934 --LSGDKPKADFQSLKLGKS---LSNPTQNKPD-----LSSLKPPKQSEE 1974
Qy 1591 GKAS 1594
Db 1975 NKPS 1978

RESULT 10
O76891 PRELIMINARY; PRT; 5327 AA.
AC O76891;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE EG:49B4.1 PROTEIN.
GN FUTSCH OR EG:49B4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louls C.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL031128; CA20006.1; .
DR Flybase; FBgn0015390; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FEFEE23A118FF38A CRC64;

Query Match 3.1%; Score 278; DB 5; Length 5327;
Best Local Similarity 19.7%; Pred. No. 3.4e-08;

Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

Qy 2 QSIMDSSAVNATEATEQNDGSRQDVLEFDLNKTPOKPKRKRKMPKVYVVEG-----K 55
Db 1986 ESID-----EAEKSESRRE-----SVAEKSLPLPSKEASRPASVAESIKDEAEK 2031
Qy 56 PKRPRKPAELPKVYVEGPKRKP-----RKAATQKVKSKETGSAKKNLKESATKK 108
Db 2032 SKEESRRESVAEKSPLPSKEASRPASVAESIKDEAEKSEESRRESAAEKSPLPSKEASR 2091
Qy 109 PANYG-----DMSNKSPEVTLKSCRKALNFLENPDGAR--OGDSE--SEIVQNSGANSF 160
Db 2092 PASVAESVKDEADSKSESRRE-----SMAESGKAQSIKGDQSPLESRSPRESVAESV 2144
Qy 161 SEIRDAIGTNGSFLDSVSQIDKTNGLGAMNQPLEVSMGNQPKLSTGAKLARQOQPDLL 220
Db 2145 KD--DPVKSEPSRRESVAGSVTADSDQSPLESKASRPESVDSVKDEAEKQESRR 2202
Qy 221 TRNQCOFFPVATONTQPFPMENQQAWMKQNLIGLFPFGNQQRMTIRNQOCLAMGNQOP 280
Db 2203 ESKTESVIPPKAKDKSPKEVLQ-----PVSMTB---TIREDA-----DQP 2240
Qy 281 MYLGTFRPALVSNQO-----LGGQGNKRPFIPLNHQTCLPAGNQLYGSPTDMHQLV 333
Db 2241 M-----KPSQAESRRESIAESIKASSPRDEKPSLASKEASRPASVAESIKYDLDKPOII 2294

```
QY 334 MSTGQQH-----GLLIKN-----QQGSLIRQQPCVPLID--QQPATPK-----GF 374
Db 2295 KDKSTEHSRESLEDKSAVTSKSVSRPLSVASDHEAAVAIEDDAKSSISPKDKSRPGF 2354
QY 375 THLNQWVATSSPGLRPHSOSQVPTTLVHVESVR-----ILNGTGTQORSRAPAYD 428
Db 2355 -----VAETVSSP-----IEEATMEFSKIEVVEKSSLALSLOGSGGKLTQDSSPY-- 2400
QY 429 SIQQDIHQGNKYLSHSIENGNGCKKALPONSLPTPTIMAKLEEARGSKROYHRAMGOTE 488
Db 2401 ----DVAEGD---FHAV-----SVSTVPTLTTPAELA-----QIGAATV 2436
QY 489 KHDLLAQIAQS--QDVERHNSSTCYEILDAAKTKIKQVQVENLHGMPEVEIETED-- 545
Db 2437 SSPLDALRTSPAPHSIRADSPACASEEIASQDKSPQVLKES--SRPANVAESKDDAA 2494
QY 546 ----PTDGARKGNKTASTISKASKGN--SSPVKKT-----AEKEKIVTKTPAKKRA 592
Db 2495 QLKSSVEDLRSPVASTEISRPASAGETASPIEEAPKDFAEFEQAEKAVLPLTIELKGNL 2554
QY 593 GRKKSVPPAHASEIQLMQPTPPKPTPLSRKPKGKRSIODSGKARGPSGELLCOISIA 652
Db 2555 PTLSSPVDVAHAS--VQPAELSKVDIEKTASSPIDEAPKSLIGSPAERPEPSAESAKDAA 2613
QY 653 ELIYRMQNLXLGDKEREQEQNAMVL-----YKGDGALVPYESKRRKPRPKVDIDDETT 705
Db 2614 ESV-----EKSKDASRPSPVVESTKADSTKGDISPSP--ESVLEGPDKDDVEKSESS 2663
QY 706 RIWLLMKGDEKEGEDEBKDKKKEKWEERVRFRGRADSFIAHMLVQDGRFRSPWKG 765
Db 2664 RPPSV-----SASITGDSTKDVSR-----PAS 2685
QY 766 VVDSVIGVFLQNVSDHLLSSAFNSLAARFPKLLSSREDERNVSVVVEDEPGCILN 825
Db 2686 VVESV-----KDEH-----DKAESRRESIAKVESVI--DEAG----- 2715
QY 826 ETPSQWQKVQHPDSMEVSGVDSGSKQELRDCSNSGIERFNEKSIQNL-----EVL 879
Db 2716 -----KSDSKSSQDSQDKDEKTLASKEASRESVSVESKDDAEKSESPEVSI 2764
QY 880 SS-----QDSFDPAIFQCGRGVSC-----SCSKSDAEFPPTTRETCTVSGTS 922
Db 2765 ASGEVPVRESKPLDSKDTSRPGSVVESVTADEKSEQQSRRESVAESVKADTKR--DGKS 2823
QY 923 QSVOTGSPNLDEICLOGNE---RPHLYEGSDVQKQETINVAQKPKDLEKTMNWKDSV 978
Db 2824 Q--EASRFSVDELKDDDEKQESRRQITGSHKAMSTMWGDESMDKADKSKESRSPESV 2881
QY 979 CFGQPRDNTNQTTSPSSYEOCATRQPHVLD--IEDFGMQGEGGLGYSMMSIS--PR--VDRV 1034
Db 2882 AESIKHENTKDEESPLGS-----RRDSVAESIKSDITKGEKSPLPKSREVSRRPESVSGSI 2935
QY 1035 KKNVPRFRFGGSGVPREFTGII--PSTPHELPGMLSGSSSAVOEHQDDTQHNOOD 1091
Db 2936 KDEAKESRRESVAESVKPESKSDATSAPPKSEHSRP-----ESVLGSLKD 2980
QY 1092 EMNKASHLOKTFDLNLSSECLRQSTK-----ONITDGLCPDRDTA-----EDVVD 1140
Db 2981 EGDKTTSRVSVADSIKDEKSLLSVQASRPESEAESLUKDAAPQSQETSRPESVETESVKD 3040
QY 1141 PLSNSSLQNLIVLSSNKKQETAVYEKTNATILREMK--GTLADGKKPTQOWDSLRKDV 1199
Db 3041 GKSPVASKEASRPASVAENAKDSADESQRPESLPQSKAGSIKDEKSPASKDEAEKSK 3100
QY 1200 ENEGQRERNKNNDSIDYEAIRASISEI--SEAIKERGNMML-----AVRIK 1247
Db 3101 E--ESRRESVAEQFPVLSKEYSRPASVAESVKDEAEKSEESPLMSKDEASRPASVAGSVK 3158
QY 1248 DFLERIVDHGIDLEWLRSPDPKA 1273
Db 3159 DEAEKSE-----ESRRESVAESK 3177
```

```
RESULT 11
Q95YM2
ID Q95YM2 PRELIMINARY; PRT: 17352 AA.
AC Q95YM2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE I-CONNECTIN.
GN I-CON.
OS Procambarus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidae; Cambaridae; Procambarus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RL sarcomeres of crayfish claw muscle.";
RL EMBO J. 20:4826-4835(2001).
DR EMBL; AB055861; BAB64297.1; -.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 2.8%; Score 253; DB 5; Length 17352;
Best local Similarity 17.9%; Pred. No. 7.9e-06;
Matches 364; Conservative 336; Mismatches 789; Indels 550; Gaps 88;

QY 6 DSSAVNAPEATEAQDNGSR-----OD---VLEFDLNLKTPQ---OKPSKRK 43
Db 11703 ESISVSQTOVHESVDDVRVAKLKEKALTILPQDSITVQEISVKEAPQIDKHKPKTEQ 11762
QY 44 RKFM-----PKVVVEGKPKRKPKP-----AELPKVVV 71
Db 11763 AKSILSPRESTAVEINEKPSVSDVVKPTEQAQKSIIPSHESLTVEVTVKETPKDIS 11822
QY 72 ECKPKRPRKAATOKYKSKETGSAKKNLKESATKKPANYGDMNKSPEVT-LKSCRKA 130
Db 11823 EVKPKSQTAKSISWE---QESISVQEVTVKDA---PGSVKDDKPKTETATSIIPSHES 11874
QY 131 LNFD---LENPGARQGDSESEIV---QNSGANSFSEIRDAIGGTNGSFILDSVSQ 180
Db 11875 LTQEVSVKKEAPGDIKDDKPKERATSIILSPQESIATQEVSV-VREA---PGSVKDSIPK 11929
QY 181 IDKNTGLGMNQPL---EVSNGNQPKLS-----TGAKLARD 214
Db 11930 KEQATSIISTHESVTVEHVKEEKPDISDKPKTEKATSIILSEQESISVEEVSKEAPG 11989
QY 215 QPDLTLRNQOCQFPVATQNTQFPMEHQOAMLOMKNOLIGFPNGQOPRMTIRNOOPCLA 274
Db 11990 SVKDLKUTEQAASIIIPSHES---LTQEVTVKEAPVEITDKKPKKEKATSIILSQEESIA 12046
QY 275 MGN---QOPMVLITPRP-----ALVSCNQQL-----GPOQNKRP-----IF 309
Db 12047 VQEVSVKEAPYSMEEDKPKAEKATSIIPSHQSLITEEVTVKEASGNSIDKKPKTEKALSI 12106
QY 310 LNHOTCLPAGNOLYGPSPTDMHOLVNSTGGQHQHLLIKNQPGSLIRGOOPCV----- 361
Db 12107 LSEQESISV-----QEVSVKEAPGSVKETKPKTEQATSVISPHESLTVEHVSUGE 12156
QY 362 --PLIDQOPATPKGFTHLNQMVATSM-----SSPG---LRHQSQOV---PTTVLH 404
Db 12157 AHQDLSQDKLTERATSVILSEQESISIQEILVADPAGSVKDIKPKSEQATSIIPSHESLT 12216
QY 405 VESVSRILNGTTGTCQSRAPAYDSLQODIHQGNKYILHSHEIS--NGNGCKKALPQNSSL 462
Db 12217 VQEVSVKESPEISDKRPKSEKATSIILSE---QESISVHESVADPAGSKMDKPKTEQ 12272
QY 463 PTPIMAKLE-----EARGSKROYHRAMGOTEKHDLNLAQIAQS--QDVERHNSSTCTVE 514
```

Db 12273 ATSVISPHESLTVQEVSVKSSSTSEISDKPKSEKATSIILSEQESISVQEIISVKDAPGSIK 12332
QY 515 YLDAAKTKTIQKVQENLHGMPPEVIEDPTDGTARKGNTASISKAGSKGNSSPVKKT 574
Db 12333 --DAKPKTEQATSIISPHESITIQEVSVKETPTDIDSRKPKTEKATSIILSEQESISVQEV 12390
QY 575 AKEKCKIVPWT-PAKGRAGRKSVPPPAHASEIQLWQPTTPKTPLSRSPKPKGKRKSI- 632
Db 12391 SVKE--APGTVETAKPKTEQATSIILPHESISIQEISVKEAPTDICDKPKPKSEKATSIIL 12447
QY 633 --ODS-----GKARG-----PSGELLQCD-----SIAEIIYRMQNLVILGDKERE 669
Db 12448 SEQESIAVQEVSVKEAPGSEIEAKPKTEYAKSSISPHESLTVQEVIVLKEASAEISDKPK 12507
QY 670 QEQNAMVLYKGDGALVYESKKRPRKVDIDDETTIWNLLMG-----KGDEK 718
Db 12508 TGRKATSIILSEQESIAVQEVSVKEAPGRVEDVKPKTEQATSIILSEHOSLTVQEVTVKGD 12567
QY 719 EGDEKDKKKEKWEERRVFRGRADSFIAHMLVQ-----GDRRFPSPKWS 765
Db 12568 EISDKPKTE-----KAISILSEQESISVQEIITVKEAPGSVKEDLKPKQAT 12614
QY 766 VVDSVIGVFLTVQNVSDHLLSSAFMSLAARP--PKLSS--SREDERNVRSVVVEDPCCI 821
Db 12615 SVISPLESLTVQEVSVKESG---DISDRPKTEKATSIIFSEQESISVQEISVKEAPGTV 12671
QY 822 LNLNEIPSWOEKVQHPDMFVSGVDGS-KEQLRDCNSGTERNFLEKSTQNLEEVLS 880
Db 12672 ADVK--PKEQATSIILSEHOSLTVQEVTVKEIPTDISD-----KKPKSEKATSIILS 12720
QY 881 QDSFDDPAIFQSCRGVSGSCSK-----SDAEFTTTCETKTVS 919
Db 12721 EQESISVHEVSVKDPAGSMKDAKPKTEQATSVISPHESLTVQEVSVREVPTETISDKPKS 12780
QY 920 GTSQSVOTGSPNLS-DEICLOGNERPHLYEGSGDVQOQET----- 958
Db 12781 EKATSIILSEQESISVQEISVK--EAP-----GSMKDAKQKTEQATSVISPHESLTVQEVSV 12834
QY 959 ----TNAQAQKPDLEKTMNNKDSVCFQPRNDTNMQTTPSSVBOCATRPHVLD---I 1010
Db 12835 KEAPTELSDRKPREKA-----TSILSEQESISVQEVSVKEAPGSV 12875
QY 1011 EDFQMGEGGLYSWMSISPR-----VDRVKNKNVPRF-----FRQGSVPREPTGOIIP 1060
Db 12876 KDLKLTKE---QATSVISPHESLTVQEVSVKEAPTEISDKPKTEQATSIILSEQESIS- 12931
QY 1061 STPHELPGMGLSGS-----SSAVQEH-----QDQTOHQOQDEMKNKASHLQKTF 1104
Db 12932 ---QEISVKGAPGLKDEKPKSEQATSIILSEHOSLTVQEVTVKEAPADISDVKPKTEKAT 12988
QY 1105 DLLNSSECLTROSSTKO---NITDGLPDRDRTAEDVVDPLSNSSLQNTLV----- 1153
Db 12989 SILSEQESISVQEISVKEAPGSMKD-AKPTEQATSVISP-HESLTVQEVSVKEVPTEIS 13046
QY 1154 -----ESNSSNKEQTAVEYKETNATILRMKGTADGKKPTSQWDSL----- 1195
Db 13047 DKPKSEKATSIILSEQESISVHEVS---VKESPGSMKDAKPKTEQATSVISPHESLTVQ 13103
QY 1196 -----RKDVEGNEGROGRKNNKNDSDYIAIRASTISEIAKE-RGNMNLAVRIKD 1248
Db 13104 VSVKESPTESLDDKPKKEKATSIILSEQESISIQEVSVKEVPESMKDLKPKTEQATSVISE 13163
QY 1249 FLERIVKHGGIDLEWLRESPPD-----KAKDYLLSIRGLGLKSVCEVRLTLHLNL 1299
Db 13164 LLSUSVQEV-----VKESPGSLSDKKPKTERATSIILSEQESISVQEIS----- 13207
QY 1300 AFPVDTNVGRIAVRMGVPLQPLPESLQLHLELVPVLESTQKFLWPLRCLCKLDORTLYEL 1359
Db 13208 -----VKAPGSVE---EVKPKTEQATSIISPHM-----SL 13235
QY 1360 HYQLITGKVKFCYSK---RPNC-NACPMRGECRHFA-----SAYASARIALPAPEERSL 1409
Db 13236 TVQEVTVKEVPADISDVKPKSEKATSIILSEQESISVQEISVKEAPGSMKDAKPKTEQ--- 13292

QY 1410 TSATIPVPPSPFPVPAIPMIPLPLEKSLASGAPSNRENCEPIIEEPASPGQECTEITE 1469
Db 13293 -ATSVISPHESLTVQEVSVREVPTETI-----SDKKPKSEKATSIILSE-----QESISVQ 13341
QY 1470 SDIEDAYNEDDPEIPIKINIQFGMTLREHMERNMELQEGDMSKALVALHPTTTTSIPT 1529
Db 13342 ISVKEA-----PESMKDAKPKTEQATSVISPH-----ESLTVQEVSVKEV-----PTEISDKK 13389
QY 1530 PKLKNISRLRTEHQ-----VVELPDSHRLLDGMDKREPDPSPY-----LLAIWT 1574
Db 13390 PKTEKATSIILSEQESISVQEVSVKEAPGSVKDLKPKTEQATSVISPHESLTVQEVSVKEA 13449
QY 1575 PGETANSAQPEQKCGKASGKCMCFDETCSECNSREANSQTVRGTLILPCRTAMRGSFP 1634
Db 13450 PTEISDKPKPKTEQ-----ATSVLSEQESI-----SIQEISVKEAPGSL-----KDEKP 13492
QY 1635 LNGTYFOVNEFLFADHES-SLKPTIDV---PRDWIWDLPRTVYFGTSVTSIFRGLSTQOI 1689
Db 13493 KSE---QATSIILSEHOSLTVQEVTVKEAPAD-ISDVKPKTEK-ATSIILSEQESISVQEI 13546
RESULT 12
Q9HCY0 PRELIMINARY; PRT; 2254 AA.
AC Q9HCY0;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR-LIKE NUCLEAR REGULATOR.
GN TFNR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100861; PubMed=11161782;
RA Kelter A.R., Herchenbach J., Wirth B.;
RT "The transcription factor-like nuclear regulator (TFNR) contains a
RT novel 55-amino-acid motif repeated nine times and maps closely to
RT SMN1.";
RL Genomics 70:315-326(2000).
DR EMBL; AJ238520; CAC04245.1; .
DR InterPro; IPR009077; DNA_Ligase.
DR InterPro; IPR001005; Myb_DNA_bind.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00395; SANT; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN_1.
SQ SEQUENCE 2254 AA; 252844 MW; F350E96F53F04CFE CRC64;
Query Match 2.7%; Score 245; DB 4; Length 2254;
Best Local Similarity 18.7%; Pred. No. 1.3e-06;
Matches 358; Conservative 250; Mismatches 722; Indels 584; Gaps 85;
QY 11 NATEATEQNDGSRD-----VLEFDLNTPOOK----PSKRRKEMPKVIVVSGKPKRK 59
Db 339 NKFREKRTNGRIDKAFQEKRPDFFFAHLLQKVLAEERKQKQSVKNHSLAEKKSTK 398
QY 60 PRKPAELPKVVVEG-----KPKRKPKRAAT----- 84
Db 399 PRKNVKKVACEGVNNPDMSNRISDTERSOKDAQTVEEESLTLSDREDAQVALEVD 458
QY 85 --QEKVSKETGSAAK---XNLKESATKK--PANVGDMNK----SPEVTLKSCRK---- 129
Db 459 LNQKRRRKQDGNAGNELGVNLLNENATVQAGPSKGEKHKNCQAIRPELKEGEGSKQML 518
QY 130 --ALNFD-----LENPGDARQDSESEIVQNSS-----GANSFS 161
Db 519 SCTQNDIGIVGFASTEKVEKRTDPILSLNNQDATSVATSESESSTDLPSFEVIGALC 578
QY 162 EIRDAIGG---TNGSFLDSVQIDKTNGL-----GAMNOPLEVSMGNQPKLSTGAKLA 212

Db 579 EVNNAEGSCIERNVLDKNNNSLEIDOTENVKPMRLGRFORP-----KPNLSRAGKSV 631
Qy 213 RDQ-QPDLTLTRNOQOPVATQNTQFMENQOAWLOKMNQLIGFPGNQOQPRMTIRNOQP 271
Db 632 LSQKTESKNSHSKSVKXNHVKKMKNLFDLIRMETTERENP-----EATVSVLGEKN 688
Qy 272 CLAMNQOPMYLIGTPRPALVSGNQGLGGPOGNK-----RPIFLN-----HOTC 315
Db 689 CLQESQ-----LKALRPVQVGRGLQPKPNAGRAERKEILISQEEIGANVEKNENESC 743
Qy 316 LPAGNQLVSGTDMHOLVMTGGQOGLLIKNOQPGSLIRQQP-----CVPLIDQQ 367
Db 744 A-----DRDTPQHMEQDSKDFEEDVILQPEKNDSFQVQDEPKVLNECLSVQENN 796
Qy 368 PATPKGFTHLQMWATSMSPGLPHSOSQVPTTYLHVESVRLNGTTCQCRSRAPAY 427
Db 797 KA-----NKLQV-----PILTRFQKPKP-----NIGRGTRREISKE 831
Qy 428 DLSQODIHQGNKYILSHBISNGCKKALPONSSLPPIMAKLEEARSGSKROYHRAMQOT 487
Db 832 EVLEKILVSGEMAAALRETVRDLDTSPEM-----VPAEINTK-EMQSDLKETGRRALSPR 885
Qy 488 EKHDNLQAQIAQSQDVERHNSSTCV-----EYLDAAKTKIKVQOE-----NLHG 534
Db 886 EKILDVIDDITMETGLKAMGREICLREKTPVIDATE--EIDKDLAAGRREISPOKNG 943
Qy 535 MPVEIETEDDPTGARKGKNTASISGASKNSSPVKTAKEKCI--VPKTPAKGR- 591
Db 944 -PEEVKPLGEVETD-----LKATGNSSPREKTPGEGTDATEIDKNLEETGRR 990
Qy 592 -----AGRKS-----VPPPAHAS-BIQLWQP-----TPKTP 618
Db 991 KISPENGPEEVKPVDEMETDLNATGRESSPREKTPVIDATEIDLEETERESVPOENG 1050
Qy 619 LSRKPKGKRKSTQDSGK-----ARGSGELLQCDISAEIYRMQNLVGLKERE--OEQN 673
Db 1051 LEEVKPLGEMETDLKATGROSFPFGKTPVEI--DAIEI-----FIDLEETEREISPOEN 1103
Qy 674 AMVLYKGDGAL---VPVESKRRKPRPV-DIDDETTRINWLLMGKDEKDEKKKE 729
Db 1104 GLEEVKPLGEMQTDLKATGREISPREKTPVIDAT-----EIDKDL 1146
Qy 730 KWEEERVFRGRADSPFIARMLVQGRDRFPSPKGSVVDVSVIGVFLFQNVSDHLSASF 789
Db 1147 E-----TGRREISP-----BENGPEEVKPVDEM 1169
Qy 790 SLAARFPKLLSSREDERNVRSVVDEPDEGCILNLNLPNSWQKVQHPSDMEVSGVDSGS 849
Db 1170 ETDLKTTGREGSSREKTR-----VIDAAEVIETDLEE-----TEREISPOENGP 1214
Qy 850 KEQLRDCSNSGIERFNFLKSIQNLLEEVLSQDSFDPALFQSGRGVSCSCSKSDAEFP 909
Db 1215 EE-----VKPVGRMETDLKEREISQRE-----KVLAFFS 1245
Qy 910 TTR-----CET-----KTVSGTSQSV-----QTGSPNL-----SDEICLOGNER 943
Db 1246 AIREKEIDLKATGKRDPIMEKVSGKMAVVEEMADLKETGKFNFRGSEEICV----- 1300
Qy 944 PHLYEGSDVQKQETTNAVQKKPDLKTMWKNDSVCGFQPRNDTNWQTTSSSYEQCATR 1003
Db 1301 ---TEEKVAELKQTKTDISPRENLEETSTR-----QTDHLMQSGSNDFSAMPS- 1349
Qy 1004 QPHVLDIEDFCMOQEGLYGWSNISPRVDRVKNKNVPRRF--PROGG-----SVPREFTGQ 1057
Db 1350 -----LDLQNTSSE-----VLSMMHTPVEKRNKESEVSHSFHFKISSQTHESDKTEVOGI 1401
Qy 1058 IIPSTPHELPMGLUGSSSAVQEHQDDTQHQNDQDMNKASHLQKTFDLDLLNSSECLTRQ 1117
Db 1402 QSPDVPEQFSDINLSKSL-----PQEQKPLEIKPAPFVRSRF----- 1438
Qy 1118 SSTQKNTDGLDRDRAED--VVDPLSNSSNLONILVSNSSNKEGTAVEYKETNATIL 1175

Db 1439 KRKPENLARAALKRETTESEKYYIEKKSEKTKMETIYMQENNEQDTLPSQHDASLMIS 1498
Qy 1176 REMKGTIADGKK-----PTSQW-----DSLKRDVEGNEGROERNKNNDSI 1216
Db 1499 RE-KDTLCHRNEEAIVLPCTQTERNLSPNSCEPKEESQAPVQKNDVSVSVGTNNVNTF 1557
Qy 1217 DYEAIRRASISESEAIKERGMNNMLAVRIKDFELERIVKDHG---GIDLEWLRESPPDKA 1273
Db 1558 QQE---MKESVIQTARQVRGLRQRPRIKRTQORQIVDKGEAKIIEGRTILPKDET 1613
Qy 1274 KDYLSTIRGLGLKSEVCRLLTLHNLAFPDVTNNVGRITAVRMGWVPLQPLPSLQHLLEL 1333
Db 1614 EK-----KVLTVSN--SOIETEIE-----VPSSAVPE----- 1638
Qy 1334 YPVLESIQKFLWPLRLKLDQRTLYELHYQLITFGKVFCTKS---RPNCNACPMRGEC- 1387
Db 1639 -----HRMYENOSQVVLVENLHVNTNETIRHENKPYVPSSAQMT 1678
Qy 1388 -RHFASAYASARLALPAPEERSLTSATIPVPPESPFPVPAIPMIELPLPLEKSLASGAPSN 1446
Db 1679 RRKFQKAKPNLGRAHSKKEPVLKVTTDQSKGKP-----EDHLLQKGSN 1725
Qy 1447 -----RNCSPITIEEPASQOECTEITESDIEDAYYNEDDPEI--PTIKLNEIQEG---M 1496
Db 1726 TQLLLKEKABELLTSLEVSARKDCVGSRESAL--AKIDAEELEVGPSSRRVGETVGDNSPS 1783
Qy 1497 TLREHMERN-----MELOEGDMSKALVALHPPTTSTPTKLNKISRLRTEHQVVELPD- 1549
Db 1784 SVVEEQYLNKLTSCPPQPLNETSYSK--IALDGKTTISTSEYERNRGERSHKAFK-PNV 1840
Qy 1550 -----SHRLLDGMDKREPDDPSPYLLAIWTPGTANSAQPPSEQCGKSGKCMCFDETC 1604
Db 1841 TRGRGSKRVGKTSKKEPRASKAMLVTLRASQEDDDADD-----FESDYE 1886
Qy 1605 ECNSLRFRANSQTVRGTLIPCTAMRGSPFLNGTYFQVNFELFADHESSLKPIDV 1658
Db 1887 E-ESYHLAPEVANKAPFVFP--VGLRSPFVSA---QIETMEELTEITVNPV 1934
RESULT 13
Q9H197 PRELIMINARY; PRT; 2187 AA.
AC Q9H197;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE TENR PROTEIN (FRAGMENT).
GN TENR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100861; PubMed=11161782;
RA Kelter A.R., Herchenbach J., Wirth B.;
RT "The transcription factor-like nuclear regulator (TENR) contains a
RT novel 55-amino-acid motif repeated nine times and maps closely to
RT SN1.";
RL Genomics 70:315-326(2000).
DR EMBL; AJ279120; CAC21448.1; JOINED.
DR EMBL; AJ279121; CAC21448.1; JOINED.
DR EMBL; AJ279122; CAC21448.1; JOINED.
DR EMBL; AJ279123; CAC21448.1; JOINED.
DR EMBL; AJ279124; CAC21448.1; JOINED.
DR EMBL; AJ279125; CAC21448.1; JOINED.
DR EMBL; AJ279126; CAC21448.1; JOINED.
DR EMBL; AJ279127; CAC21448.1; JOINED.
DR EMBL; AJ279128; CAC21448.1; JOINED.
DR EMBL; AJ279129; CAC21448.1; JOINED.
DR EMBL; AJ279130; CAC21448.1; JOINED.
DR EMBL; AJ279131; CAC21448.1; JOINED.
DR EMBL; AJ279132; CAC21448.1; JOINED.


```
DR EMBL; AJ279133; CAC21448.1; JOINED.
DR EMBL; AJ279134; CAC21448.1; JOINED.
DR EMBL; AJ279135; CAC21448.1; JOINED.
DR EMBL; AJ279136; CAC21448.1; JOINED.
DR EMBL; AJ279137; CAC21448.1; JOINED.
DR EMBL; AJ279138; CAC21448.1; JOINED.
DR EMBL; AJ279139; CAC21448.1; JOINED.
DR EMBL; AJ279140; CAC21448.1; JOINED.
DR EMBL; AJ279141; CAC21448.1; JOINED.
DR EMBL; AJ279142; CAC21448.1; JOINED.
DR EMBL; AJ279143; CAC21448.1; JOINED.
DR EMBL; AJ279144; CAC21448.1; JOINED.
DR EMBL; AJ279145; CAC21448.1; JOINED.
DR EMBL; AJ279146; CAC21448.1; JOINED.
DR EMBL; AJ279147; CAC21448.1; JOINED.
DR EMBL; AJ279148; CAC21448.1; JOINED.
DR EMBL; AJ279149; CAC21448.1; JOINED.
DR EMBL; AJ279150; CAC21448.1; JOINED.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001005; Myb_DNA_bind.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00395; SANT; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN_1.
FT' NON_TER 2187
SQ SEQUENCE 2187 AA; 245726 MW; 2FB083F5F82AFB55 CRC64;

Query Match          2.7%; Score 241; DB 4; Length 2187;
Best Local Similarity 18.4%; Pred. No. 2.2e-06;
Matches 345; Conservative 254; Mismatches 685; Indels 594; Gaps 82;

QY 2 Q5IMSSAVNATEATEQNDGSRQDVFLEFDLNTPOQKPSKRRK-----FMPKYV 51
Db 430 RSQDAQTVEESLTLSREDAEQVALEVDLN---QKKRRKKQDGANELGVNNLENAT 485
QY 52 VEGKP-----KRPRKPAELPKVVVGKPKRKAATQ-----EKVSKETGSA 96
Db 486 VOAGPSKGEKHKHKKCOAIRPE-LKEGEGSKEQMLSCTQNIIDGIVGFASTKEVE-KRTDPI 543
QY 97 KKKNLKESATKPPANVGMSNK---SPEVTLK-----SCRKALNFLEN----- 137
Db 544 LSLNQDQATSVATESSESTSDLSFEVIGIRALCEVNNABGSCIEERNVDLKNNSLEID 603
QY 138 -----PGDAR-----QGDSESEIVQNSSGA-----NSF 160
Db 604 QTENVKPLMRGRFRQPKPNLSRAGKKSVLQCKTESESKNSHKSITSVEKNHVEKDKMNTL 663
QY 161 SEIRDAIGTNGSFLDSVSQIDKTNGLAMNQ-----PLEVSMGNQPKLSTGAKLARDQ 215
Db 664 DILRMETTERENPEAETVSDLGKKNLQEGSQLKALRPVQVRGLQPKPNAGKAERKE 723
QY 216 ----OPDL-----LTRNQOCQFPVATONTQTFPMENQOAWLQMKNLIGFP-----FGNQO 261
Db 724 ILSQSEIGANVEKNEESC-----ADRTQPMEDQSRKDFEEDVILQPKNDSFQNVQ 779
QY 262 P-----RMTIRNOQCPCLAMNQOPMYLIGTPRPALVSGNQOGLGGPQGNK 305
Db 780 PDEPKVNBCLSVQENNRANKLNQVPILRTREFQPKPNIGRG-----TGRREISSKEEVL 834
QY 306 RPIFLNHQ--TCLPAGNQLYGSPTDHLQVLVMTGCGQOGLLILKNQOPGSLIRGQPCVPL 363
Db 835 EKILVSGEMAAALRETVRLDTSKPMVPAEINTKEMQSDL---KETGRAISPREKILDV 891
QY 364 IDQOPATPKGFTHLNQWATSSPSGLRPHSOSQVPTTYLHVESVSRILNGTTGTCQQR 423
Db 892 IDDTMETGLKAMREI-----CLREK 914
QY 424 AP-----AYDSLOQDIHQGNKYILSHEIS-NGNGCKKALPQNSLSLPTPIMAKLEEARGSKR 478
Db 915 TPEVIDATEEIDKLEAGR---REISPOKNGPEEVKPL-GEVETDLKATGNES-SPRE 968
QY 479 QYHRAMQOTEKHIDLNLAQIAQSQDVERHNSSTCVEYLDAAKTKTIQKVYVQENLHGMPE 538
Db 915 TPEVIDATEEIDKLEAGR---REISPOKNGPEEVKPL-GEVETDLKATGNES-SPRE 968
```

```
Db 969 KTPECTDATEEIDKNL-BETGRRKISPRENG-----PEE 1001
QY 539 VIEIEDDDPTDGARKGKNTASISKGASKGNSPVVKTAEKEKCIVPKTPAKKGAGRKKS 598
Db 1002 YKPVDEMETDLNATOR-----ESSPREKTP-----VIDAT----- 1032
QY 599 PPPAHASEIQLWQP-----TPPKTPLSRSPKPKGKRKSQDSGK---ARGPSGELLQDSI 651
Db 1033 -----EEDLEETEREVSPQENGLVEVKPLCEMETDLKATGRDSFPRGKTEVI--DAI 1084
QY 652 AEIIYRMQNLILGDKERE--QEQNAMVLYKGDGAL---VPYESKKRKPAPKV-DIDDETT 705
Db 1085 EEI-----EIDLEETEREISPOENGLVEVKPLGEMOTDLKATGREISPREKTPVIDAT- 1138
QY 706 RIWNLMLGKGEKGEDEKDKKKKEKWEERVFRGRADSFIAARMHLVQGDGRFRSPWKS 765
Db 1139 -----EIDKDLDEE-----TGREISP----- 1155
QY 766 VVDSVIGVFLTONVSHLSSSAFMSLAARFPKPLSSRDERNRVSVVDEPEGCILNLN 825
Db 1156 -----EENGPEEVKPVDEMETDLKTTGREGSSREKTR-----VIDAAEVETDLE 1201
QY 826 EIPSQEKKVQHPSDMEVSGVDSGSKEQLRDCNSGIERENFLEKSIQNLLEEVLSSQDSF 885
Db 1202 E-----TEREISPOENGP-----VKPVGMETDLKEIREISORE--- 1238
QY 886 DPAIFQSGRGVSGSCSDAEFTTR-----CET-----KTVSGTSQSV----- 925
Db 1239 -----KVLAFFSAIREKEIDLKGTGRDIPIMEKVSCKMAVVEEMAD 1281
QY 926 --QTGSPNL-----SDEICLQGNRPHLYEGSDGVQKQETNNVAAKKPDLEKTMNKDVC 979
Db 1282 LKETCKENFRERGSEICV-----TEEKVAELQTKGTDISPRENELEETSTSR--- 1330
QY 980 FCQPRNDTNWQTPSSSYEQCATRQPHVLDIEDFMQEGEGLYSWMSSISPRVDRVKNKV 1039
Db 1331 ----QTDTHLMOSGNDFSAMPS-----LDIQNISSE-----VLSMMHTPVEEKRNSEK 1377
QY 1040 PRRF--FRGG---SVPREFTGQIIPSTPHELPGMGLSGSSSAVQEHQDDTHHQQDDM 1093
Db 1378 SSHFSHFKISSOTHESDKTEVQIOSDPVPEQFSDINLSKSL-----PQEKPLBI 1428
QY 1094 NKASHLQKTFDLNLSSEECTLRQSTKQNTDGCPLRDRDAED--VVDPLSNNSSLQNI 1151
Db 1429 KPAPFVRSR-----KRPKPNLARAALKRETTESEKVIYEKKSETKKMETI 1474
QY 1152 LVESNSSKQETAVYKETNATILREMGTLADGKK-----PTSQW----- 1192
Db 1475 VMOENNEQTDLPSQHDEASLMSR-KDTLGHRRNEEAIVLPCTQTERNLSPNSCEPK 1533
QY 1193 DSLRKDVEGNEGRQERNKNMDSIDYEAIRRASISEIFAERGMNNMLAVRIKDFLER 1252
Db 1534 ESQSAPVKNDVSVSGTNNVNTFQOE-----MKESVIQTARQVRGLQRPRIKRTGOR 1589
QY 1253 IVKDHG---GIDLEWLRSPDKAKDYLLSIRGLGLKSECVRLTLHLNLAFPVDTNVGR 1309
Db 1590 QIVDKGEAKGIKEGRTILPKDTEK-----KVLTVSN--SQIETEIE- 1630
QY 1310 IAVRMGWTPQLPESLQLHLELYPVLESIOKFLWPLRCLKDQRTLHELHQLITFGVK 1369
Db 1631 -----VPSSAVPE-----HRMYENQSQVVLVENL 1654
QY 1370 FCTKS-----RPNACPMRGEC--RHEASAYASARLALPAPEERSLTSATIPVPPEFP 1422
Db 1655 HVNKNETIRHENKYPVFPSSAQMTRRKQOKAPNLGRHAUSKEEVPVLEKVTVDQSKGPK 1714
QY 1423 PVAIPMIELPLPLEKSLASGAPSN-----RENCEPIIEEPASPGQETEITEIDEDAY 1477
Db 1715 -----EDHLLQKGSANTQLLKEKAELLTSLEVSARKDCVGSRESAL--AKI 1759
QY 1478 NEDPDEI--PTKLNTEQFG-----MTLRHEMERN-----MELQEGDMSKALVALHPTTTS 1526
Db 1760 DAEELEVGPSSRRVGETVGDNSPSSVVEEYLNKLTSCPOPLNETSYSK--IALDGKTTI 1817
```



```
Qy 1214 DSIDYAIRRASISEISAIKERGMNMLAVRIKDFLRIYVDHG-GIDLEWLR----- 1266
Db 2459 DSIINVEYKSTKQKI-----DPVASVPVKRPRPVKKAQALELAGLKPIQTL 2509
Qy 1267 -----ESPDPK-----AKDYLLSLRGLGKSV 1289
Db 2510 GHTDESADKGTQMPNCNLOPQVDSFQSPRRPRTRRGKVEADEEPSAVR-----KTVS 2565
Qy 1290 CVRLTLHLNLAFFVDTNVG-----RIAVRMGW-VPLQPLPE 1324
Db 2566 TSR-QTMRSRKVPETGNGTQVSKASIKOTLDTVAKVTGSRRLRTHKGGWSTLLKLLGD 2624
Qy 1325 SIQL-----HLELVPVLESIOK-----FLWP-RLCKLDQRTLYELHYQLITFGKVC 1371
Db 2625 SKEITQISDHSEKLAHDSILKSTQOQKPDVSKPLRTCRRLRASKEVPKEVLDTRDHA 2684
Qy 1372 T-KSRPNCNACPMRCEHRHFAVASARLAL-----PAPERSLTSATIPVPES 1420
Db 2685 TLOSKSNPLSPKRSARDGSIVTRALRSAPKQEADEKVPVPEKKAASKRVSPE- 2743
Qy 1421 FPPVAIPMIELPLPLEKSLAS-----GAPSNRENCEPIIEFPASPGQECTETESDIEDA 1475
Db 2744 --PVKMKHLKI---VSNKLESVEEQVSTMTEEMAKRENVPVTDQ-----NS 2787
Qy 1476 YNEDPDEIPTKLNIEQFMTLREHMERNMELQEDMSKALVALHPTTSIPTPKL--- 1532
Db 2788 RYRK-----KTVKQ-----PREKFDAS 2805
Qy 1533 -----KNISRLRTEHOVYELPDSHRLLDGMKREPDPSPYLLAIWTPGETANSAPPP 1585
Db 2806 AENVGIKNEKTMKTASOETEL-----GNPD-----GAKKSTSR- 2840
Qy 1586 EQCGKASGKCMCFDETSECNSREANSQTVRGTLTI--PCRTAMRGSGFP 1634
Db 2841 -----GOVSGK-----RTC-----LRS-----RGTEMPQCEAEKTSKP 2871

RESULT 15
Qy 9520
ID Q9520 PRELIMINARY; PRT; 2701 AA.
AC Q9520;
DT 01-NOV-1999 (TreeBLrel. 12, Created)
DT 01-NOV-1999 (TreeBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TreeBLrel. 14, Last annotation update)
DE HYPOTHETICAL 295.8 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rhodes S., Huckle E.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL096857; CAB51071.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2701 AA; 295828 MW; D2043FF9840ADECD CRC64;

Query Match 2.6%; Score 239; DB 4; Length 2701;
Best Local Similarity 18.6%; Pred No. 4.1e-06;
Matches 352; Conservative 225; Mismatches 675; Indels 640; Gaps 87;

Qy 14 EATEQNDGSRQVLEFDLNTKPTQQPKRKRKFMKPVVVEGPKRKRKPAELPKVVVEG 73
Db 502 EIREREREKEREREKEKEQEREREK-----DRERQEREREK-----EQ 549
Qy 74 KPRKRPRAATQEKVKYKTSAGKKNLKESATKPAWGDMSKNSPEVTLKSKRKALNF 133
Db 550 EKOREMERKEKEKELEKEREKELEK-----MKEQEKEREKEKELEKEE 598
Qy 134 DLENGDARGDSESEIYONSSGANSFSFEIRDAI--GGTNGSFLLDSVQIDKTNGLGAM 190
Db 599 KIE-----PREPNLE--PMVEKESENSCKNEEPVETRODSNRSEKEATPVVHET----- 647
```

```
Qy 191 NOPLEVSMNGNOPDKLSTGAKLARDQOQPDLLT---RNQOQFPVATQNTQFPMENQOAWLQ 247
Db 648 -----EPESGSP-----RPAVLSGYKQFKSLPPRFQRQOQEQMKQO--WQO 689
Qy 248 MKNQLIGFPFNGQOQPRMTIRNOQPCLANCNOQPMYLLIGTPPALVSGNQOQGGPOGKRP 307
Db 690 QOQGVLPQTVPSQPSSTVPPP-----HRPLVQPMQHP-----OHLASMGFDRW 737
Qy 308 IFLNHOTCLPAGNOLYSGPTDMHQLVMSTGGQOQGLLTKNQOQPSLIRGQOQPCVPLI--D 365
Db 738 LMM-----QSYMDDR-----MMSG--RPAMDPIPIHGM-----PPPLMRDR 774
Qy 366 QOQATP---KGFTHLNMV---ATSMSSPGL-----RPHSQSQVPTTYLHVSVSRILN 413
Db 775 QMEGSPNSSEFELIARSARDHAISLSEPRMLWGSQDPYHAEPOQATTPKATEEPEDV-- 832
Qy 414 GTTGTQCSRA-----PAYDSLOQDIHQGNKYLLSHEI----- 446
Db 833 -----RSEALDOEQITAAYSVEHNOLEAHPKADFIRESSEAOVQKFLSRSEVEDVRPH 885
Qy 447 ---SNGNGCKKALPONSSSLPTP---IMAKLEEARGSKROYHGRAMQOTEKHDLNLAQOIA 499
Db 886 HTDANNQOACFEA-PDQKTLAPQOEERISAVESQPSRKRSYSHGSHHTQKQDEQSEPSA 944
Qy 500 QSQDVHRNSCTCVELD---AAKTKIQVVOENLHGMPEVIEIEDDPTDGARKGN 555
Db 945 GIPKV---TSRCIDSKEPIERPEKPKKEGFIRSESGPKPKVYKSKSETRWGPRPSSN 1000
Qy 556 TASTSGASKNSPVKK-----TAEK---EKCIVPKTPAKKGRAGR 594
Db 1001 RREEVNDPRVRRSRGIPKPVLRDMKEEREQKEKEGAEKVTKVVV--RPEKTEKDL 1058
Qy 595 KKSVPAPAHASEIOLWQPTPKTPLRSRSPKPKG---GRKSTQDSGKARGSGELLCCDS 650
Db 1059 PPPPPPOPPAPIQOSVPPPIQPEAEKFPSTETATLAQKPSQTEKLEPLEVSVQVPEA 1118
Qy 651 IAEIIRMONLYLGKEREQEQNAMVLYKGDGALVYVESKRRKPRP---KVIDDETT--- 705
Db 1119 VKTV---NOQTMAAPVVKEQKPEKISK-----DLVIERPRPDSRPVAKKESTLPP 1167
Qy 706 -RIWNLMLGKDEKDEKDEKKEKKEWEEERRVPRGRAD-----SFIARMHLVQ 754
Db 1168 RTYW-----KEARERDWFQDQYGRGRGEYYSRGRSYRGSYGRGGRGR 1212
Qy 755 GDRRESPW---KGSVDSVIGVFLTONVSDHLSSAFMSLAAREPPKLSLSRREDERVNR 810
Db 1213 GHTRDYPOYRNKPRAEHIPSGPLRQRESETRSESSDFEVV---PKRRQRQSETD 1268
Qy 811 SVVVE---DPEGCILNLNEIPSQEKVQHPHSGVDSGSKQELRDCSNSGIERFNFL 867
Db 1269 SEIHESASDKDS--LSKGLPKREERPENKKPVK-----PHSSFKPDNHV 1311
Qy 868 EKSQNLLEEVLSSQDSFDPAL-----FOSCGR-VGSCSCSKSDAEFFPTR 912
Db 1312 RIDNRLLEKPYVRDDDKAKPGFLPKGEPTRRGRGTFRRGDRDGGGRPSRSTLRRPAYR 1371
Qy 913 -CETKTVSGTSQVQSGSPNLSDE--ICLOGNERPHLYEGSDGVQKQE----- 957
Db 1372 DNQWNPPOSEVPKPEDGEPFRRHEOFIFIAADKRPKPEKFERKFDAPRPRRQRTPRPQ 1431
Qy 958 -----TTNVAQ-----KKPOLEKTMNKD 976
Db 1432 DKPPFRRLREREAASKNEVVAVPTGNTVNNVAQEPVNTLGDISGNKTPDL-SNONSSD 1490
Qy 977 SVCFQGPNDNTNWQT-TPSSSYEQCATQPHV---LDTEFGMQGEGGLGYSWMSISPR-- 1030
Db 1491 QA-----NEEWETASESDFNERRERDEKKNADLNAAQTVVGENVLPKREIAKRSF 1543
Qy 1031 -----VDRV-----KNKNVPRFRQSGSVPREFTGQIIP--STPHELPGMGL 1071
Db 1544 SSQRPVDRQNRGNNGPPKSGRNFSGPRNRRSG---PPSKSGKRGKPPDDQAGTTGVDL 1600
```

Qy	1072	SGSSAVQEHODDTQHNODEMANKASHLOKTFDLN	SSBECLTRQSTQNTQNTDGLCLR	1131
Db	1601	INGSSA	-----HHQGVPNGTG--QK-----NSKDTGKKREDPK-----	PG 1635
Qy	1132	DRTAEDVDPLSN	-----NSSLONILVESNSSKEOTAVYKETNATILREMKGLTADGKKP	1188
Db	1636	PKPKPKYDALSOFDLNNYASVVIID	-----DHIPEVTVIDPQSN--LNDGFTFVYSKKQ	1689
Qy	1189	TSQWDSLRDKVEGNEGRQER	-----NKNMNSIDYEAIRRASISEIAKERGNMNM	1241
Db	1690	-----QKRLQDEERRKKEEQVQWNNKANE	-----KGRSQTSK	1724
Qy	1242	LAVRIKDFLERIVKHGGIDLEWLRESPPDKAKDYLLSIRGLGKSVCEVRLTLHLNLAFL	1301	
Db	1725	L-----	-----PPRFACKQATGQ-----QAQSSASVPLASAPL	1754
Qy	1302	PVDNTNGRIAVRMGWVP	-----LOPLPESLOHLLELYPVLESIQKFLWPLRCLKDQRTLIEL	1359
Db	1755	PPSTSASVPASTSAPLPATLTPVPASTSA	-----PVPAS-----	1788
Qy	1360	HYOLITFGVKFTKSRPNACNPMGECRHFASAYASARLALPAPERSTUTSATIPVPE	1419	
Db	1789	-----TLAPVLASTAP-VPASPL	-----APVSASASVSVPASTSAAAITSSAPA	1835
Qy	1420	SFP-----PVAIDMI	-----ELPPLPLEKSLAS-GAPSNRENCPIITEEPASPQCEC	1464
Db	1836	SAPATPILASVSTFASVTILASAPI-LASALASTAPT	-----APASASSPAP-----	1886
Qy	1465	TEITESDIEDAYNEDDPDEIPTKINIEQFGMTLREHMERNMELQEGDMSKALVALHPTT	1524	
Db	1887	-VITAPTI-----PASAPT-	-----ASVPLAPAS	1908
Qy	1525	TSIPTPKLKNISRLRTEHOVYELPDSHRLLDGMDKREDDPSPVLLAIWTFGETANSAPQ	1584	
Db	1909	ASAPAPAPTVS-----	-----APNPAPPAP-----AQQAQTHKPVQN	1942
Qy	1585	PEQKCGKGASGMCFDETCECN	-----SLREANSQTIVRGFTLLPICRTAMKSGFPLNGTY	1639
Db	1943	PLQ-----	-----TTSQSSKQPPSIRLPSNAQTPNGT-----	1969
Qy	1640	FQWNELPADHESLKPIDVPRDW	-----IWD	1665
Db	1970	-----DYVASGKSQIOTPPQSHGTLTAELWD	1993	

Search completed: September 27, 2002, 13:13:22
Job time: 1115 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 10:04:11 ; Search time 238.43 Seconds
(without alignments)
805.463 Million cell updates/sec

Title: US-09-840-743-2
Perfect score: 9089
Sequence: 1 MQSIMSSAVNATEQND.....PRPLMARLHPFASKLNKNT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

```
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246.5	2.7	3256	21 AAY50976	Human cell cycle p
2	244.5	2.7	2724	22 ABG20119	Novel human diagno
3	237	2.6	2819	22 AAB35408	Human 07CG27 gene
4	232	2.6	2290	22 ABB60066	Drosophila melanog
5	230.5	2.5	2063	22 AAM78710	Human protein seq
6	230.5	2.5	2063	22 AAM40064	Human polypeptide
7	230	2.5	3263	22 ABB67210	Drosophila melanog
8	225.5	2.5	2063	22 ABG17147	Novel human diagno
9	225.5	2.5	2519	22 ABB16636	Novel human diagno
10	225.5	2.5	2759	22 ABG17148	Novel human diagno
11	222.5	2.4	5447	22 AAU14697	Novel bone marrow

12	222	2.4	3257	22 ABB67502	Drosophila melanog
13	221.5	2.4	5533	22 ABB65772	Drosophila melanog
14	221.5	2.4	5560	22 ABB71160	Drosophila melanog
15	216	2.4	5373	22 AAU14603	Novel bone marrow
16	214.5	2.4	2362	22 ABB58396	Drosophila melanog
17	213	2.3	2263	22 AAM79000	Human protein seq
18	211.5	2.3	2439	22 ABB68896	Drosophila melanog
19	210.5	2.3	1552	22 ABB71764	Drosophila melanog
20	210.5	2.3	1805	13 AAR27204	Rat nestin. Rattu
21	210.5	2.3	1805	15 AAR60126	Rat nestin protein
22	210	2.3	1852	22 ABB67170	Drosophila melanog
23	210	2.3	2016	22 ABB71487	Drosophila melanog
24	208.5	2.3	2432	21 AAY85565	Human homologue of
25	207	2.3	1637	22 AAM00916	Human bone marrow
26	207	2.3	2951	22 ABB60291	Drosophila melanog
27	205.5	2.3	3111	22 ABB60327	Drosophila melanog
28	205	2.3	1982	22 ABG16404	Novel human diagno
29	205	2.3	1982	22 ABB19656	Novel human diagno
30	205	2.3	6815	22 ABB68811	Drosophila melanog
31	204.5	2.2	2194	22 AAM40114	Human polypeptide
32	201	2.2	1703	22 ABB66223	Drosophila melanog
33	201	2.2	3238	22 ABB71715	Drosophila melanog
34	201	2.2	5024	22 AAG82935	S. epidermidis ope
35	200.5	2.2	2703	22 ABB60074	Drosophila melanog
36	198.5	2.2	1920	22 ABB64441	Drosophila melanog
37	198.5	2.2	1976	22 ABB69419	Drosophila melanog
38	198	2.2	1984	22 ABB61060	Drosophila melanog
39	197.5	2.2	2400	22 ABG20278	Novel human diagno
40	197.5	2.2	3201	22 ABB62899	Drosophila melanog
41	196.5	2.2	1795	22 ABG21018	Novel human diagno
42	196	2.2	2189	22 AAM79017	Human protein seq
43	196	2.2	2285	22 ABB63057	Drosophila melanog
44	195.5	2.2	2161	22 AAM78959	Human protein seq
45	195.5	2.2	2523	22 AAU03503	Human protein kina

ALIGNMENTS

```
RESULT 1
AAY50976
ID AAY50976 standard; Protein; 3256 AA.
XX AC
XX AAY50976;
XX AC
XX 10-MAR-2000 (first entry)
XX DE
XX Human cell cycle protein Ki-67.
XX
XX Cell cycle protein; Ki-67; therapy; cell proliferation; allergy;
KW tumor treatment; autoimmune disease; scar formation; inflammation;
KW rheumatic disease; transplantation.
XX
XX Homo sapiens.
XX
XX DE19822954-AA1.
XX
XX PD 25-NOV-1999.
XX
XX PF 22-MAY-1998; 98DE-1022954.
XX
XX PR 22-MAY-1998; 98DE-1022954.
XX
XX (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
XX Flad H, Gerdes J, Boehle A, Deinert I;
XX WPI; 2000-039964/04.
XX N-PSDB; AA43872.
XX
XX Ki-67 gene antisense oligonucleotide -
XX
XX Disclosure; Page 19-29; 36pp; German.
PS
```

XX	This invention describes a novel oligoribo- or oligodeoxyribonucleotide, characterized in that, it hybridizes to mRNA that encodes protein Ki-67 at a physiologically acceptable salt concentration. The oligoribo- or oligodeoxyribonucleotide which is complementary to Ki-67, a protein active at all stages of the cell cycle except G ₀ , is useful for therapy of illnesses with increased cell proliferation and particularly for treatment of tumors, autoimmune diseases, scar formation, inflammation, allergy, rheumatic diseases and defence against transplantation. This sequence represents the human cell cycle protein Ki-67 which is described in the method of the invention.			
XX	Sequence	3256 AA;		
QY	Query Match	2.7%; Score 246.5; DB 21; Length 3256;		
CC	Best Local Similarity	18.8%; Pred. No 6e-09;		
CC	Matches	362; Conservative 266; Mismatches 768; Indels 527; Gaps 94;		
QY	22	SRDVLFEFLNKTPOQPKSRKRKFKMPKVVEGK-----PKRKPR--KPABELPKVVVEGKP 75		
Db	1496	sqdpvdtptsskpskrsr-----kdvveeffalrkrtpsagkamhtpkpavsgk 1549		
QY	76	KRPKRKAATQEKVKSTGTGSAKKNL---KESA-----TKKPANVGDMSNK 118		
Db	1550	niyafmgtpvgkldltenltgskrrlqtpekaqaledlagfkelfqtrghteesmctndk 1609		
QY	119	SPEYTLKSKALNFDLENPCDARQGDSESEIVQNSSGANSFSIRDAIG---GTNG--- 172		
Db	1610	takvacksqpdl---knpasskr-----lktslgkvkveellavglkltsgett 1660		
QY	173	-----SFLDVSQSI--DKTNGLGAMNOPLEVSMGNO--PDKLSTGAKLARDQ 215		
Db	1661	hthteptgdgskmfafmespkqildsaasltgskrrlqtpkgsevpbedlagfielf--q 1718		
QY	216	OPDLLTRNOQCFFVATQNTQFPMENQ-----AWLQMKNOLIGFPFGNQ--QPRMTIR--- 267		
Db	1719	tp-----shktesmnekttkvsyrasqpdldvdtptskpqpksrslkad 1763		
QY	268	NOQPCLAMGNQPM--YLIGTPRPALYSNGOQLGGPGQGNRPFLNHQTCLPAGNQLYGS 325		
Db	1764	teeeflafrktpsgagkamhtpkpav-----geekdi-----atflgt 1801		
QY	326	PTDMHQLVMSTGGQGHLLIKNQPGSL--IRG-----QPCV--PLIDQOPATPKGFTH 376		
Db	1802	pvqklqpgnlpgsnrllqtrkekaqaleltgfrelftptcndptadek-----tt 1854		
QY	377	LNQWATSMSPGLRPHSPQSQVPTTYLHVESVS-----RILNGTTGTCQSRAPAYDSL 430		
Db	1855	kkilckspqdpadtptntkqprksrskadvveeflafrkltspagkamhtpkaaqve- 1913		
QY	431	QODIHQ--GNKYILSHEISNGCGKALPONSSLPTPIMAKLEARGSKROY-----HRAM 484		
Db	1914	ekdintfvgtpvekldlignlpgsrr--pqt---pkakaledlagfkelftptpghtee 1969		
QY	485	GOTEKHDNLAAQQAQSOQDVHNSSTCEYLDAAKTKIKQVQENLHGMPEVIEIED 544		
Db	1970	sntddkitesvckspqdpvktptss-----kqrlkslsgkv-----gvkeevlpv-- 2015		
QY	545	DPTDGARKGNKTASISKA-----SKGNSSPVKKTAEKKECIV-----PKTPA 587		
Db	2016	-----gkltqtsqgkttqthretagdgksikafkesakqmdpanvgtgmerwprtpk 2067		
QY	588	KKGR-----AGRKKSVPVPAHASE-----IOLWQPTP-----PKTPLSR 621		
Db	2068	eaqsladlagfkelfqtpdhteestddtktkackspppsmdtptsttrirtpktpigk 2127		
QY	622	-----SKPKG---KGRKSIOQSGK-----ARGPSGE 644		
Db	2128	rdiveelsalkqltqthtdkvpvgdedkvinvfretakqklqdpaaavtsgskrqprtpkgk 2187		
QY	645	LLCQDSIAEIIYRMONLYLGDREQEONAMVLYKGD-----GALVPVESKKRKPAPKVD 699		
Db	2188	aqpledlagkelfqtpvctdkptthektkiacrspqdpvgtpgtifkpskrsrskad 2247		
QY	700	IDDETRIRNLLMGKDEKE-----GDEKDKKKKEWEEERVFGRGRADSFARMHL-- 752		
Db	2248	veeeslalkrtpsvgkamdtpkpagdekdmk-----afmgtp---vqkldlpg 2294		
QY	753	-VOGDRRFPW-----KGSVDSVIG--VFLTQNVSOHLSSSAPMSLAARFP----- 796		
Db	2295	nlpgskr---wpqtpekaqaledlagfkelfqtpgtddkptdtktkiackspqdpvd 2351		
QY	797	PKLSSSREDERNVRSVVVEDEPGCIILNLNIPQWQKVQHPSDMEVSGVDSGKEQLRDC 856		
Db	2352	tpastkqprkrlrkaadv--eeflalkrtpsgakamdtp-----kpavsd 2397		
QY	857	SNSGIERFNFLEKSIONLE-----EEVLSODSDFDPAIFOSCG----- 894		
Db	2398	knln-----tfvetpvqkldlignlpqskrqtpqtkaealedlvf--keifqpghtee 2452		
QY	895	-----RVGSCSKSDAE--PPTTRCE-----TKTVSGTSQSQTG 928		
Db	2453	smtddkitevsckspqsfktrsskqrlkiplkvdmkeeplavsklrltsgettqth 2512		
QY	929	SPNLSDBEICLQ--NERP--HLYEGSDV--QKQETTNNVAQKKPDLEKTMNWKDSVCFQCP 984		
Db	2513	teptgdsksikafkespkqildpaasvtsgrqrltrrkekaraledlvdfke--lfsapg 2570		
QY	985	NDTNWQTTPSSSVEOCATROPHVLDIEDFGMOGEGGLGSNMWISPRVDRVKNKNVPRRF 1044		
Db	2571	hteemtdkntktpcksppttdt-----atckrcpkr-----pkrev 2612		
QY	1045	ROGGSVPREFTGQIIPST--PHELPGMLSGSSSAVOE-----HQDDTQHNOODENKK 1095		
Db	2613	keelsaverltqsgqsthtkhepasdegikvlkqrakkkpnpveeepssrripapkek 2672		
QY	1096	ASHLOKT--FLDLLNS--BECLTROSSYKQNTDCLPRDRTAEDVWDPLSNSSSLON 1150		
Db	2673	aqpledlagftelsetsqhtsagtagkatk-----ipcespplevvdtastkhrhlr 2725		
QY	1151	ILVESNSNKQTAVEYKQETNATI-----LRMKGTADGKKPTQOWDSLRLKRDVEGNE 1203		
Db	2726	trvqkvqvepsavkftqtsgettdadkepagedkgika--lkesaktppaasvtsgr 2784		
QY	1204	GRQERNKNMDSIDYAIRRASISEIPAIRKERNMNNMLAVRIKDFLERIVKDHGGIDLE 1263		
Db	2785	rrprapresaqae----dlagfdpaaghteesmtdtktkipckespeledatss-- 2838		
QY	1264	WRESPPDKA-----KDYLLSIRGLKSVCEVRLLTILHNLAFVDTVNVGRIAVRMGWVP 1318		
Db	2839	--krprtradrkvekeellavglktsge-----tchtkepvgegkgtkafk----- 2886		
QY	1319	LOPLPESLQLHLELYPVLESIQKFLWPRLCKLDQRTLYEL-----HYQLITFG 1367		
Db	2887	--qakrn-----vdaedvigrrq--prapkekaqplediasfqelsqtpghteelang 2937		
QY	1368	KVFCTKSPNCNACPMRGCEHRHFASAYASARALPAPER---SLTSATIPVPPSPFPVP 1424		
Db	2938	aadstfslap--kqtpdsgk-----plkisrivrakpvepvvgdvvstrdpvksgskant 2989		
QY	1425	ALPMIELPLPLEK---SLAGAPSNRENCEP---IIEE-PASPGQECTEITESDIEDAY 1476		
Db	2990	slp-----plpfkrsggkgdsvgtgkrlrmpapegiveelpaskkqr-----vapra 3037		
QY	1477	YNEDPDEIPTTKLNTIEQFQMTLREHMERNMELQEDMSKALVALHPTTTSIPTPK----- 1531		
Db	3038	rgkseepvimmkrsr---tsakriepaeelnsndm--ktnkeehklqdsvenkgsir 3092		
QY	1532	LKNISRLRTEHOVEL--PDSHRLLDGMDKREPDDPSPYLLAIWTGETA----- 1579		
Db	3093	srrqdkaeqitvefvlaerieinrnekpkpmktspe--mdiqnppddgarkpirdkvt 3151		
QY	1580	-----NSAQPPEQKCGGKASGKWCDF-----ETCSECNLSREANSQTVR 1618		
Db	3152	nkrcirsarqnessqpkvaeesggqksakvlmqnqkgkgeagnsdmcrlsrstktksqpa 3211		


```
Db 1455 reaasksnvavptngtnvnaqepvntlgdisgnktpdl-sngnssdga-----ne 1506
QY 988 NWQJ-TPSSSYEQATRQPHV-----LDIEFGWQEGGLGYSWMSISPR-----VDRV-- 1034
Db 1507 ewetasesdfnerredkdnadnaqtavkvgenvlppkrelakrsfssqrpvdqr 1566
QY 1035 -----KKNVPRFRFGSGVPRFTGQIIPSTPHE--LPGMGLSGSSSAVOEHQ 1082
Db 1567 rgungppksgrnfgsnrrsg---ppkskrgpfddhpagttgvdlingssa----- 1618
QY 1083 DDTQHNOQEMNKASHLQKTFLLDLNLSSECLTRQSTQNTDGCCLPRDRTAEDVVDPL 1142
Db 1619 ----hqeegvpngagg-----knsdctgkredpk-----pgpkpkpkvdal 1658
QY 1143 S-----NNSLOWILVESNKEQTAVEKFTNATILREMKGTADGKKPTSQWDSLRK 1197
Db 1659 sqvdlhnyasv---viid---dhpevtviedpgsn---luddgftevvskq-----qk 1703
QY 1198 DVBEGNEGRQBER-----NKNNDMSIDYEAIRRASISEISEAIKERGMNMNMLAVRIKDEL 1250
Db 1704 rlqdeerrkkeeqiavwnkknane-----kgrsqtskl----- 1737
QY 1251 ERIVKHGGIDLEWLRESPPDKAKYLLSTRGLGLKSVECVRLTLTLHNAFPVDTNVGRI 1310
Db 1738 -----pprfakkqatgig-----qagssasvpplasaplbpstsaas-- 1773
QY 1311 AVRMGWVPLQ---PLPESLQHLLELYPVLESIQKFLWPLRLCKLDQRTLYELHYQLITFG 1367
Db 1774 -----vpastapilpatl-----pvpastsgsv-----stltis 1803
QY 1368 KVFTKSRPNACNAPRGECRHFASAYASARLALPA--PEERSLTS-----ATIPVPPEFP 1422
Db 1804 pvwaspsap-vpasplap-----vsasasvpsastsaaitssasapasaptila 1857
QY 1423 PVAIP-----MIEPLPLEKSLAGAPSNRENCEPIIEEPASPGOECTEITESDIEDA 1475
Db 1858 svtpasvtnpasasipirasalastsaptp---apaaasspaap-----vitapti--- 1905
QY 1476 YYNEDPDEITIKLNTBQFCMLREHMERNMELQEGDMSKALVALHPTTTTSIPTPLKNI 1535
Db 1906 -----pasapt-----asvphatasasapaptpvsapnapp-- 1939
QY 1536 SRLRTEHQVYELPDSHRLLDGMBKRPDDPSYLLAIWTPGET-----ANSAPPEQKCG 1590
Db 1940 apatqgqthk-pvqnpqlgtsgskqppbsiirlpsaqtngtdyvasgksigtqgsh-- 1996
QY 1591 GKASGKWCFTCECSNLSREANSOTVRGTLIIPCRTAMRGSFPLNCTYFQVNELFADHE 1650
Db 1997 gtitaefmwdn-----kvaphqavindill-----rklg 2025
QY 1651 SSLKPIDVPRDWIWDLPRTVYFTSVTSIFRG-----LSTEQIOFCFWKGFVCV 1700
Db 2026 splvpppppsvawnkplts--fysaps--egakngqesgleigttdiqf----gapas 2077
QY 1701 RGFQKT-----RAPRPLMARLHFP-ASKLKNK 1728
Db 2078 ngnevenvpliseksadkipekqeqrkqragpikaqk 2116
RESULT 3
AAB35408
ID AAB35408 standard; Protein; 2819 AA.
XX
AC AAB35408;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human 07CG27 gene protein.
XX
KW Human; 07CG27 gene; chromosome 1; HPC1 region; prostate cancer;
XX
KW oncogene.
```

```
OS Homo'sapiens.
XX WO200116291-A2.
PN
XX 08-MAR-2001.
PD
XX 25-AUG-2000; 2000WO-US23291.
PF
XX 27-AUG-1999; 99US-0151049.
PR
XX (MYRI-) MYRIAD GENETICS INC.
XX (HOSP-) HOSPITAL FOR SICK CHILDREN.
PA
XX Tavitigian SV, Swedlund B, Simard J, Rommens JM;
PI WPI; 2001-226682/23.
XX N-PSDB; AAF28060.
DR
XX Novel human prostate cancer marker gene termed as 07CG27 gene, useful
PT for screening mutations in the gene in diagnosis of a predisposition to
PT cancer -
XX
XX Claim 1; Page 91-99; 99pp; English.
PS
XX The present invention provides the protein and coding sequences of the
CC human 07CG27 oncogene. This gene is found at the HPC1 region of
CC chromosome 1. The sequences can be used in the diagnosis and
CC identification of treatments for prostate cancer. The present sequence is
CC the 07CG27 protein.
XX
XX Sequence 2819 AA;
QY 14 EATEQNDGSRQDVFLEFDLNTPOOKFSKRRKFKMPKVVVEGKPKRPRKPAELPKVVVEG 73
Db 504 eelrererekererekekeqekerek-----dreqqekekerek-----eq 551
QY 74 KPRKPRKKAATQVKSKETGSAKKNLKESATKPKANVGDMSNKSPEVTLKSCRKALNF 133
Db 552 ekremerkerkekelekerkekelqk-----mkeqekcekelekerkelee 600
QY 134 DLENPCDARQDSESEIVQNSGANSFSEIRDAI---GGTNGSFLDSVSDIDTNGILGAM 190
Db 601 kie---prepnle-pmveqsesncnkeepvfrqdsnrsekeatpvvhet----- 649
QY 191 NQPLEVSMGNQPKLSTGAKLARQDQPDLLT---RNOCCQFPVATONTQFPMENQQAALQ 247
Db 650 ----epesgsgp-----tpavisgyfkqkqlppfrqrgqeqmkqgq-wqg 691
QY 248 MKNQLI-----GFPEGNQOP-----RMTIRNQOPCLAMGNOQPMYLIG---TPR----- 288
Db 692 qggqgvilqvtvpsqssstvpppphrplyqmqp-----hpqhlasmgfdprwlmmgs 744
QY 289 ---PALVSGNQQLGGPQGQGNKRPIFLNHQTCLP-----AGNQLYGSPTDMHQLV-MSTGGQ 339
Db 745 ymvprrmmsgrpamdip-----pi---hpgmiipkplmrrdmeqsgpnssesfehiansar 796
QY 340 QHGLLIKNOQPGSLIRGQQPCVPLIDQOPATPKGFTHLNOMVATSMSSPGLRPHSQSOVP 399
Db 797 dhaislepr---mlwgsdpyphaepqgattpkateepdv---rseaal---dqeqit 846
QY 400 TTYLHVESVSRILLNGTTGTQCRSRAPAYDSLQDDIHQGNKYILSHEI-----SNG 449
Db 847 aay-----svhe-----nql Leahpkadfireseaqvqkflrsvedvrphtdang 894
QY 450 NGCKKALPONSSSLPTP---IMAKLEEARGSKROYHRAMGOTKEKHDNLAAQOIAOSQDVER 506
Db 895 sacfea-pdqktisapceerlisavesqsrksrvshgsnhtqkpdqsrseasagipkv-- 951
```

Query Match 2.6%; Score 237; DB 22; Length 2819;

Best Local Similarity 18.6%; Pred. No. 2.6e-08;

Matches 350; Conservative 224; Mismatches 685; Indels 626; Gaps 89;

Qy 507 HNSSTCVYLD-----AAKTKIQKVQENLHGMPEVIEIEDPTDGARKGNKTASISKG 562
Db 952 --tsrcidskepieerpeekpkkegfirsesgpkpekysksetwgrprpsnnrreevnd 1009
Qy 563 ASKGNSSPVKK-----TAEK--EKCIvPKTPAKKGRAGRKSVPPP 601
Db 1010 rpvrirsgipkpvrlrmdkeereqrkekegaekvkvvv--kpektekklppppppp 1067
Qy 602 AHASBIQLWQTPPKTPPLSRSPKPKGK---GRKSTQDSKGARGSGELLCODSAEIIYR 657
Db 1068 qppapiqpsvppplqpeaekfpstetatlakpsqdekteplevstvqevavktv--- 1124
Qy 658 MQLNYLGDKEREQEONAMVLYKGDCALVPYESKRRKPRP--KVIDDDTT-----RIWNLL 711
Db 1125 nqqlmaapvvkeekqpekvisk-----dlvierpdrpavkkestlpprttyw--- 1173
Qy 712 MGKGDEKEDEKDKKEKWEERRVFRGRAD-----SFIARHMLVQGDREPS 761
Db 1174 -----kearerdwfpdggrygrgeyysrgysrgygrgrgrgrgrhtrdyp 1221
Qy 762 W----KGSVVDVIGVFLTONVSDHLSSAFMSLAARPPPKLSSRREDENRVRSVVVE-- 815
Db 1222 qyrndkpraehipsplqrseesetrsessafevv---pkrrrgsetatdseihsa 1277
Qy 816 -DPEGCIILNLEIPSWQEKVQHPDMEVSGVDGSKGEOLRDCSNSGIERFNFLEKSIQNL 874
Db 1278 sdkds--lskgklpkreerpenkpvk-----phsxfkpdnhvridnrl 1320
Qy 875 EEEVLSQDSFDPAI-----FQSCGR-VGSCSCSKSDAEFPTR-CETKTV 918
Db 1321 ekpyvrdkdkagfipkgeptrrgrgtrgrgdrpgrpsrptlrrpayrdnqwnpr 1380
Qy 919 SGTSGVGTGSPNLSD--ICLQNGRPHLYEGSGDVQKQE----- 957
Db 1381 qsevpkpedgeprhrhefipiaadkrppkferkfdparerrqrtrprqrkdppfr 1440
Qy 958 -----TTNVAQ-----KKDLEKTMNMKNDVCFQGP 983
Db 1441 rlreeraasknevvavptngtvnnvagepvtngidsgnktptdl-snqnsdga----- 1494
Qy 984 RNDTNWQF-TPSSSYEOCATQPHV---LDIEDFMQEGGLGYSMWSISPR-----VD 1032
Db 1495 --neewetasessdfrerredknaadnqvtvkvngenvilppkrejakrsfssqrpyd 1552
Qy 1033 RV-----KKNVPRFRFGGSGVPREFTGQIIP--STPHELFGMGLGSSAV 1078
Db 1553 rqnrrngnppksgnrgspnerrsg--ppsksgkrgpfdgpagtvgvdlngssa- 1608
Qy 1079 QEHQDDTOHQDDEMKNASHLOKTFELDLNSEECLTRQSTKQKNITDGCPLPRDRTAEDV 1138
Db 1609 -----hhqegvpngtg--qk-----nskdstgkredpk-----pgpkpkpek 1644
Qy 1139 VDPLSN---NSLQNLIVESNSSNKEQTAVEYKFNATILREMKCTLADGKKPTSQWDSL 1195
Db 1645 vdaIsqfIdnnyasvviId----dhpevfieidpdsn-----lnddgftevsvkxq----- 1691
Qy 1196 RKDVEGNEGROER-----NKNMNDSDIYEAIRRASISEIAIKERGMMNMLAVRIKD 1248
Db 1692 qkrlqdearrkkeeqvlgvwnkknane-----kgrsqtskl----- 1727
Qy 1249 FLERIVKHGGIDLEWLRESPPDKAKOYLLSIRGLGLKSECVRLTLHNLAFVVDTVNG 1308
Db 1728 -----pprfakkqatgiq-----qaqssasvpplasaplpsstas 1763
Qy 1309 RIAVRMGWVP--LQPLPSLQLHLELIPVLESIQKFLWPRCLKLDQRTLYELHYQLITF 1366
Db 1764 vpastsaplpatltpvpastsa-----pvpas-----tl 1792
Qy 1367 GKVFCTKSRPCNACPMRGECRHFASAVASARLALPAPEERSLTSATIPVPPESFP---- 1422
Db 1793 apvlstasap-vpaspl-----apvasasvasvpastsaaitssasapapapt 1844
Qy 1423 -----PVAIPMI----ELPLPLEKSLAS-GAPSNRENCEPIIEPSPGQECTEITESD 1471

Db 1845 ilasvstpsvtillasasipil-lasalastsapt-----apaasspaap-----vitapt 1894
Qy 1472 IEDAYYNEDPDEITIKLNIQFOGTMLEHMERNMELQEGDMSKALVALHPTTTTIPTPK 1531
Db 1895 i-----pasapt-----asvplapasasapapa 1917
Qy 1532 LKNISRLRTEHQVVELPDSHRLLDGMDKREPDPSPYLLAIWTPGETANSAPPPQKCGG 1591
Db 1918 ptpvs-----apnpappap-----aqtaqthkpvqnplq----- 1947
Qy 1592 KASGKMCDEFCTSECN-----SLREANSQTVRGTLILPCRTAMRGSPFLNGTYFOVNELF 1646
Db 1948 -----ttsqsskqpppsirllpsaqtpngt----- 1971
Qy 1647 ADHESSLKPIDVPRDW-----IWD 1665
Db 1972 -dyvasgksigtqshgtltaelwd 1995
RESULT 4
ABB60066
ID ABB60066 standard; Protein; 2290 AA.
XX AC ABB60066;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 6990.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL04169.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 6990; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2290 AA;

Query Match 2.6%; Score 232; DB 22; Length 2290;

Best Local Similarity 19.18, Pred. No. 4.6e-08;

Matches 356; Conservative 208; Mismatches 668; Indels 636; Gaps 82;

```
Qy 40 SKRKKKPKVVEGCKPKRKPKA---ELPKVV--VEGPKPKRPK--KAATQEKVKRSKE 92
Db 8 nkrirk-----etelpppppvpagsslapsvaaiesppmpkrtkvaqrfagqg 62
Qy 93 TGSARKKKNLKESAT-----KKPANVGDMSNKSP-EVTLKSCRKALNF----- 133
Db 63 sapsssiayaaaaaagsgpstgsadnqpdiellpnkcpkqdfllfclfrgtra 122
Qy 134 -----DLNPGDARGDSESEIVQNSGANSFSEIRDAIGTNGSFLDSVQIDKTNGLG 188
Db 123 lpahldfInqgskdseasstgannnnarkpnnsraslaguk-----kkrjrp 171
Qy 189 AMNQPLEYS---MGNOPDKLSTGAKLARDQOPDLLTRNQOCQFPVATQNTQFPMENQQA 244
Db 172 akgatckvddvwapldpctekpaekpasd----- 201
Qy 245 WLOMKNOLIGFPFGNQOPRMTIRNQOPLAMGNQOPMYLIGTPR---PALVSGNQQLGGP 301
Db 202 ---knnli---aaeapklpsrntpvlp-----gavkravevldgnrr--ga 241
Qy 302 OGKRPFLNHOQLCPAGNQLYGSEPTDMHQLVMSTGGQOGLLLTKNQOPGSLIRGOQPCV 361
Db 242 rg--rgagishdktppaa-----lkdnkrrsnrastkesk 275
Qy 362 PLI-----DOOPATPKGTHLMQVATSMSSPGLRPHSOSOVPTTYLHVESVSR--I 411
Db 276 pviddnsedldegeaddedefsahdatrsngeevkadspvkaepckrprsaakav 335
Qy 412 LNTGTG--TCQRSAPAYDSLOQDIHQGNKYILSHIEISNGCKKALPONSSLPTPIMAKL 470
Db 336 epaegqvtlkdapkvrvqr---espifipsessgrmrtraffepvkvpph----- 387
Qy 471 BEARGS-----KQYHRAMGQTEKHDLNLAQOIAQSQDVERHNSSTCYEYLDAAKT 522
Db 388 keepahdpdkdepkkqpetppsktek-eqepthkqikeak--ektnssi-----yakea 438
Qy 523 KIQKVQVQNLHG---MPEVEIEIDPDGARKGKNTASISKGASKGNSPVKKTAERKE 579
Db 439 kqqlengdsllhgavakgkksinvfssddeqlakskllkga-kq-aspkakesvee- 495
Qy 580 CIVPTPAKKGAGRKKSVPVPPAHASEIQLWQTPPKTPLRSKPKGKRKSIODSGAR 639
Db 496 ---pbggkrgkagkkaedapadeedl-----appsskkvatnarg-srksak--skas 544
Qy 640 GPSGELLQDS-----IAELIYR---MQNLXLGD-----KERQOE 671
Db 545 edsaeleprsqrparktkaaaymglghklqladdeedlsmssfpdipnvkemekm 604
Qy 672 QNAMLVYKGDGALVPYES-----KKRKPRLPK---VDIDDETT----- 705
Db 605 eneiKrnaagklnvpeastvlsagkstrkprpkdsptvtaetrnespsppeaeek 664
Qy 706 -----RIWNLKMGKDEKGEDEKKKKEKKEWEEERRVFRG 741
Db 665 pppparrrgppkqkvnapghkpriegmlvkkgslakmseqgakpkak---sehsksvdsd 722
Qy 742 RADSFIAHMLVQGDPR---FS-----PWKGSVVD 768
Db 723 eeedflineinetkrkileksfdsddepilaikvpavavkekellppkmaspppkllvpvp 782
Qy 769 SVIGV-----PLTONVSDHLSS-SAFMSLAARFPKP 798
Db 783 spisvptpavavapalavpsilptstgttqltmpltkatatslgsmmpfisiqasypl 842
Qy 799 LSSREDERNVRSVVVEDEPGCILNL-----NEIPSWQEKVQH----- 836
Db 843 ktmekkp-----pyptskilnvpatyalpgsasgavsfqktsstylpqaasphyhsy 893
Qy 837 -----PSDMEVSGVDSKSEQLRDCS-----NSGIERE-----NFKLESIQ 872
Db 837 ----- 872
```

```
Db 894 vrpptpthqfggagapsktpsgsgsplkyqtpttpttpppieayqcpkinpnfl---tp 950
Qy 873 NLEBEVL--SSQDFDPAIFQSCGRVGSQSCSKSDAEPTTTRCETKTIVSGTSQS----- 924
Db 951 kyersplprsstsfp-----spspvkfqptsqgttatvgtlpsaatkspl 995
Qy 925 -----VOTGSPNLSDICLOGNERPHLYEGSGDVQKQETTNVAQKPKDLEKT 971
Db 996 apppppsvsvqigtsslnlmslaatvnaap-----mpvaastaaaahgttvatt 1047
Qy 972 MNWKDSVCFQGRNDTNWQITTPSSSYEQCATRQPHVLIEDFGMGEGGLGYSWMSISPRV 1031
Db 1048 v-----tatstppahnsaat-----ggag--ansdpik 1076
Qy 1032 DRVKNKNVRRFRFGGSGVPREFTGQIIPSTPHEPLPGMGLSGSSSAVQEHQDDTQHNOOD 1091
Db 1077 dei-----gsilaqat--lmpskeesgkifgiavslaqsgpdkctct--- 1118
Qy 1092 EMNKASHLQKTF-----DLNSEEICITROSSTKQNTIDGCLPRDRTAEDVVD 1140
Db 1119 -lgcgsihkpvlgpvtvtegyfgdqlsskerrkakvnmtheqikwliecssnpdeiqd 1177
Qy 1141 PLSNNSSLONTLVESNNSKQETAWEYKETNATILREMGKTLADGKKPTQSDSLRKDVE 1200
Db 1178 dladd--fdslrpgqstpppplr--dkelsasfsssskntrgdlgkeasawakgpsik 1234
Qy 1201 G-----NEGRQERNKNMDSIDYEATRRASISISEAIRKERNMMLAVRIKDFL 1250
Db 1235 atpvlvtpnrrkelheqadskdcaldyd--ksstpvnlqtkits---neslav----- 1283
Qy 1251 ERIYKHGGIDLEWLRESPPDKAKDYLLSIRGLG-----LKSVECVRLLLTLHNLAFPV- 1303
Db 1284 ekkvndrkg-----kesakaaakaaatqtrstaatpbtptissttptpaslt-psksstpt 1336
Qy 1304 -----DTNVGRIAVRMGWVPLQLPESLQLHLLLELVPLESIQKFLWPRLCK 1350
Db 1337 ppavkqkaekakrnatagggatalaspaaaptpanpk-----rlpv 1377
Qy 1351 LDQRTLYELHYQLITFGVFCTKSRPNCNACPMRGECRHFASAYASAR----- 1398
Db 1378 ynqknaqaqqq-----aetkpsan--ppsgagtkresvayafgkddesgksgrnr 1426
Qy 1399 -----LALDAP-----EERSLTSATIPVPPESPFPVPAIPMIELPL-PLEKSLASG----- 1442
Db 1427 rtdpsvpapalvpnlalers-----ptkkrasaaaaatavqtlstptenckiegkpsk 1480
Qy 1443 APSNRENCPIIEEPASPQCEITEESDIEDAY----- 1477
Db 1481 aptgrgakkqggqapapappvpeasgdsdaegatfiplqgavvgsgdggigqvavklg 1540
Qy 1478 NEDPDEIPTIKLNIEQFGMTLREHMERNE-----LOEGDMSKALVALHPTTTSIPTPKL 1532
Db 1541 regpdg-pnqkv-vmqatlvtkagmdtnskplpeslnelvktilhaasndaaatttsl 1598
Qy 1533 KNISRLRT 1540
Db 1599 kslpkast 1606
```

RESULT 5

AAM78710

ID AAM78710 standard; Protein; 2063 AA.

XX

XX AAM78710;

XX AC

DT 06-NOV-2001 (first entry)

XX

DE Human protein SEQ ID NO 1372.

XX

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
tissue growth factor; immunomodulatory; cancer; leukaemia;
nervous system disorder; arthritis; inflammation.

QY	1237	GMNMLAVRIKDFLERIVKDHGGIDLEWLRSPDPKADKYLLSTRGLGLSKAVECVRLTLT	1239
Db	1499	gapnv-----tikppgltdle--vtpp-----vvsgealkkasvip--tl	1534
QY	1297	HNLAFVDTNVGRIARVWGVPQLPLPESLQHLLELYPVLESTOKFLWPRCLKLDORTL	1356
Db	1535	qdlsssksepsns-----nlphusnelcslvhpelsevssnvapsipvmrsv	1583
QY	1357	YE-----LHYQLITFGVKFCTKSPNCNACPMRGECHRFASAYASARLALPAPEERSLT	1410
Db	1584	sssststppnqit--vfvst-snp-----ittsantsaalpthlqsalml	1625
QY	1411	SATIPVPFESFPVVAIPMIETLPLPLEKSLAGAPSNRENCPIIEEPASPGQECTEITES	1470
Db	1626	stvtvmpnag-----skvmvsegsgaagsnarpqfitpvf-----inss	1664
QY	1471	DIEDAYNEDDEIPIKLNIE-----QFGMTLREHMERNNEQEGDMSKALVALHPTT	1524
Db	1665	sliqvngksgqetipaaplitnsgimppsvavvgvlhlpqnk-----fssapvppnals	1719
QY	1525	TSIPTPKLKN-----ISRLRTEHQVVEIP-----DSHRLLDGMKREPPDPSPYL---	1569
Db	1720	ss-papniqtgrplvlssratpqlpspctssvvpshppvqvkvkelnpdeaspqvnts	1778
QY	1570	-----LAIWTPGETANSAQPPEQKCG-GRAS--GKMCFFDETCE--CNS	1608
Db	1779	adqntlpsqsttmvspilitnsgsgsgnrrspvssskgkvgkvgilltkackkvts	1838
QY	1609	LREANSQ 1615	
Db	1839	lekgeeq 1845	
RESULT	6		
AAAM40064			
ID	AAAM40064	standard; Protein; 2063 AA.	
AC	AAAM40064;		
XX			
XX			
DT	22-OCT-2001	(first entry)	
XX			
DE	Human polypeptide	SEQ ID NO 3209.	
XX			
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			

DR	WFI: 2001-442253/47.
XX	N-PSDB: AAI59220.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
PS	Example 5; SEQ ID NO 3209; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 2063 AA;
	Query Match 2.5%; Score 230.5; DB 22; Length 2063;
	Best Local Similarity 18.0%; Pred. No. 5.le-08;
	Matches 326; Conservative 225; Mismatches 645; Indels 611; Gaps 80;
QY	52 VEGPKRRP-----RKPAELPKVVVEGPKRKRKAATOEKV----- 88 :: : : : : :
Db	407 lggpsrvptlqqhlcnkspasspsfsgssptvntqgqgmprppqnplpgg 466
QY	89 -----KSKETG-SAKKKNLKSATKKPANV 112 : : : : : : :
Db	467 fqpqvsspgnrmvggnvpbnfmvmqgqpnqgpqlhpglggmprklrpfgsagqanp 526
QY	113 GDMSNKSPEVTILKSCRKA-----LNFDELPNGDARQDSEISEIVQSSGANSFSE---- 162 : : : : : : :
Db	527 nfmqgvvsttatipngsagpqllqanqvqhaggagapqpqq-mqvhgppnmqpslm 585
QY	163 ----IRDAIGGTSGSFSDYSQIDTKTNGLGAMNO--PLEVSMGNOPDKLTGAKLIARDQQ 216 : : : : : : :
Db	586 gihgmnnmqgatsg----ypqvnlsmnmgppqgppsqqlmgbmhqilvpsfgqmvgvqg- 639
QY	217 PLLTRNQOCFFPVATONTQTPMENQQAWLQMKNOLIGFP----FGNQOQPRMTIRNQOPC 272 : : : :
Db	640 -----gtlnpqnpmillsraqllmpqggmmvnppbsnlgpspqmt----- 678
QY	273 LAMGNQOPWYLIGTPRPALVSGNQQLGGP-----OGNKRPFLFNH 312 : : : : : : :
Db	679 -----pbkqmlsgqpgqmmaphngmnpqggvlllqqpmieqimtgmgknkgfntqn 732
QY	313 QT-CLPAGNQLYGSP-TDMHQVLVMSTGGQHGLLIKNOQP-----GSTLRGQQ 358 : : : : : : :
Db	733 qsnvmppgaqlmrqtpnmgqnmvftgmsqgmllpqgqpvnnspqymglggqvlrppg 792
QY	359 PCVPILIDQOAPPKGFTHLNMQVATSMSGPLRHFSQSQVPTYLYHVESVRILNTGTGT 418 : : : : : : :
Db	793 ps-plumaqghgpa-tttanndvslsqmmpdvsiaqtmv----- 830
QY	419 CORSRAPAYDSLQODIHQGNKYILSHETS-----NGN-----GCKKALPNIS--SLP 463 : : : : : : :
Db	831 -----phvgamqnsasgn-hfsghgmsfnafsgapngnqmscggpnpgfvnkadvlt 883
QY	464 TPIMAKLEEARSKROYHRAMQOTEKHDLNLAQQIAQSQDVVERHNSSCTVEYLDAAKTK 523 : : : : : : :
Db	884 splvnlqlsdlsaqhi-----gvnnkqmntnanpkpkkkppr 921
QY	524 IQKVVOENLH---GMPPVEIEEDPTDGARKGNKTASTISKASKGNSSPVKKTAEKC 580

Db 148 atpiekkpapakleakvvlksikterdggimslatleqiagktekaipwitmekl 207
Qy 129 KAL-----NFDLE-----NPGDARQGDSESEIVQNSSGANSFSEIRDAIGGTNGSFID 176
Db 208 kavesvqqnlkfddevylqleg-qieteqqlpqagq-----ve 247
Qy 177 SVSQIDKTNGLGAMNOPLEVSMGNQDPKXISTGAKIARDQOPLLTRNQ-OCOPPPVATQNT 235
Db 248 qvrtkeigrilksmesveImemtldqldltqgnakdlipwkemrqqlksvgrvtqld 307
Qy 236 QPFMENQAWLQMKNLQIGFPFCNQPRWTI-----RNOQPCLAMGNOQ 279
Db 308 kfkieveirhlqaggaiteeytqttaetvnmideskgsiskvlrrideq-lyededs 365
Qy 280 PMT---LIGT-----PRPALVSGNQQLGGPQGNKRPIFLNHQTCLPAGNQL 322
Db 366 niyqkfittedvnmhvsarekleaqrlireqgavnwrgqqrp-----q 412
Qy 323 YGSP-TDMHQLVMSTGGQHQHLLIKNQOQSLIRGQ-QPCVPLIDQOPATPKGFTH---L 377
Db 413 --qpltsvedvisqtsrqkly---ggqsflieeaqrqgvvedsgmmsleeyehqkii 467
Qy 378 NOMV---ATSMSPGLRPHSOSOV-PTTYLHVE-----SVSRILNGTGTGCRSR---A 424
Db 468 nrttqqaefswrp-repqkfiqvdstllhlqerhdtqeqqllqqpvmwdrgrkkp dq 526
Qy 425 PAYDSLQDD-----IHQGNKYILSHE-----TSNGNGCKKALPQNSSL- 462
Db 527 pqyvqqeqvrvkeefvekpkyteemhdleptieqppvpvmwergkkkppqeqkte 586
Qy 463 -----PPTMAKLEARGSRQYHRAMGQTEKHDNLNAQIAQ-SODVERHNSSTCV 513
Db 587 eahdelveptpv---qqpevpvmwergkkkvaqgetvlsqevvqtsvveq----- 635
Qy 514 EYLDAAKTKIOKVQENLHGMPEVTEIEDDPTDGARKGNTASTISKGASKNSSPVKK 573
Db 636 qiweektktavrrvipprepqkvqvtltptrprkeavkaeelqklpr-strvpvq 694
Qy 574 TAEEKKCI-----VPKTPAKKGRAGRKSVPPPAHASEIOLMOTPPKT----- 617
Db 695 pveaeqkayeeatdelteepipqpvmwergkkkpkpqqeevtei-----pktleia 747
Qy 618 --PLSRSKPK-----GGRKSIQ--DSGKARGPSGELLCOCSIAFIIRMQNLY 662
Db 748 vdtleeepkptepqpqvliwargqkpkpdkqelpklsleavdtieedlik----- 802
Qy 663 LGDEREQEQNAMVLYKGDGALVPYSGKKRPRP-----KVDIDDETTR----- 706
Db 803 -----pvqpeqpvl-----werkkkppqdpqdvleekldvaptktyekavdvlpde 848
Qy 707 -----IWNLLMGKGDEKDEKD-----KKKEKWWEERRVRFRGR-- 742
Db 849 pkveekpepvlwqrgkkkpksepteevhpdvdaqietvkvkedemiveekrriktkrp 908
Qy 743 --ADSFARMHLVGDRFRFPKWSKSVDSVTCVFLTQNVSDHLSSAFMSLAARFPKLS 800
Db 909 kstkevtelfeedpeeeispeeevpqkevi-----eeieeiveekrrlkktkk--pklt 961
Qy 801 SSRDERNVRSVVVEDPGCTILNLNETPSMOKVQHPHSDMEVSGVDSGSKQLRDCNSG 860
Db 962 qqvteeetpheeilikeseevveqeeiveekkkvk-----kvkpkptvaeqqlk----- 1010
Qy 861 IERFNFLKSIQNLLEEYLSQDSFDPFAIFQSGCRVSGSCSKSDAEPTTRCTKTTSVG 920
Db 1011 -----eeelpt-eetveeeetaedqqlvve-----eskkvkk 1041
Qy 921 TSQSVQTCSPNLSDBICLOGNERP-----HLYEGSGDV-----QK 955
Db 1042 vkk--ptgtvktvdveelpgeevpveevpvedvapeeeileeeqeeivdqddeiqeqk 1099
Qy 956 QETTNVAKPKDLEKTMWKSVCGRPNNDTNWQTPSSSYEQCATRQPHVLDIEDFGM 1015
Db 1100 rkvvkakkpkktktkt-----eieieedqpeee-----vleeiige 1136

Qy 1016 QGEGGLGYSMWISPRVDRVKNKKNVPRFRFQCGSVPRFTGQIIPSTPHELPCMGLSGSS 1075
Db 1137 qee-----lterqgrkvksikkpkkv----- 1156
Qy 1076 SAVQEHQDDTOHNOQDEMKNKASHLOKTFID-----LLNSSECLTRQSTKON 1123
Db 1157 -vtektvdteqepkeesqaeekvetteepkpkpapeakveqekislkpaprkg 1215
Qy 1124 ITDGCLPRDRTADVDDPLSNNSLONILVESNSSNKE--QTAVEYKE---TNAYILRE 1177
Db 1216 l-----lpekeqveevl-----lkpvkklvavseaeqpetefevfafaittedildv 1266
Qy 1178 MGTADGKPKPTSDMSLRKDVEGNEGRNKNNDSDIYEAIRRASISEISE-----A 1232
Db 1267 tkkrvkk-kkptk-----vaaaesteepaeet-eeefeeatqpeevqveepq 1317
Qy 1233 IKERGMNNLAVRIKDFLERIVKDHGGIDLEWLRESPPPKAKDYLLSIRGLGLKSECVR 1292
Db 1318 vkevaderktapkpkrkeeliekveevalk--rvtrpkkelpqeatieevrktqg--- 1372
Qy 1293 LLTLHLNLAFPVDTNVGRIAVRMGVPLQ-----p 1321
Db 1373 -----rtsikpveevkleevdlhvkekedeivgeekrtrkvkpkphedipeip 1421
Qy 1322 LPESLQHLLE-----LYPVLESIOKFLWPRCLKDORTLY 1357
Db 1422 daeptqlaeaeihielekqkpeedqpvpwkrqekqkqveevleekkw----- 1470
Qy 1358 ELHYQLITFGKVCTKSRPNCNACPMRGECRHPASAYASARLALPAP-----EERSLTSAT 1413
Db 1471 -----sgrrrripqeqqveevqik-----pipskpieeqgkpeka 1504
Qy 1414 IPVp---PESFPPVPAIPMIEL-PLPL-----EKSILASGAPSNRENCEP 1452
Db 1505 ipdpqlvpeekpeseeeeeleelepkpedkkpkpkakkekxxxpkkkatpsvdevsee 1564
Qy 1453 IIE---EPASPGQECTEITESDIEDAYINED--PDE--IPTIKLINIEQFGMTLREHMERN 1505
Db 1565 vaepfdepiaeeadeveempvddvkvvavsedvlpjeevvpt-----eetpeakqkakhkrt 1620
Qy 1506 MELQEGDM 1513
Db 1621 krlikeasv 1628
RESULT 8
ABG17147
ID ABG17147 standard; Protein: 2063 AA.
XX
AC ABG17147;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17138.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
DD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI

Db 1626 stvtvmpnag-----skvmvsegsgaaqsnarpqfitpvf-----inss 1664

Qy 1471 DIEDAYINEDPDEIPTIKLNI-----QFCMTLREHMERMEIQEGDMSKALVALHPTT 1524

Db 1665 siiqvmkgsqstipaapltnslmpsvavvqpllipqnik-----fssapvppnals 1719

Qy 1525 TSITPPKLKN-----ISRLRTEHQVYELP-----DSHRLLDGMDKREDDPSPYL--- 1569

Db 1720 ss-papniqtgrpvlssratpqlpspctssvpvshppvqqvkelnpdeaspqvnts 1778

Qy 1570 -----LAIWTPGETANSAQPPQKCG-GKAS--GKMCDFETCSE-CNS 1608

Db 1779 adqntlpsqsttmvplntnspgssgnrrspvsskgkgkvdkigqlltckackkvtgs 1838

Qy 1609 LREANSQ 1615

Db 1839 lekgeeq 1845

RESULT 9

ID ABG16636

XX ABG16636 standard; Protein; 2519 AA.

AC ABG16636;

XX 18-FEB-2002 (first entry)

DT Novel human diagnostic protein #16627.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

KW Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX N-P5DB; AAS80823.

DR New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 46995; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (ii) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (ii) and its binding partners are useful in medical

CC imaging of sites expressing (II). (i) and (ii) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 2519 AA;

Query Match 2.5%; Score 225.5; DB 22; Length 2519;

Best Local Similarity 18.1%; Pred. No. 1.8e-07;

Matches 372; Conservative 287; Mismatches 787; Indels 611; Gaps 90;

Qy 25 DVLEFDLNKTPQPKSRKRKPKFMPKV-----VVEGKPKRKRKPAELPKVVVEGKPKRKR 80

Db 117 nllecnldqelklfvshrsharfspvgqklhrsefxfetvvlqpfllmedsgsgtevr 176

Qy 81 KAATOEKVKSK-----ETGSAKKKLKE-----SATKKPANVG 113

Db 177 lmit-daarhklvltgqcfentgelllqsgsfqnfieftdqeigellstthpanka 235

Qy 114 DMSNKSPVTLKSCRKALNFDLENPGD-ARQOSESEIVQNSGANSFSEIRD----- 165

Db 236 sltlfcpe---egdwksnldrhnlqdfiniklnsasilpemelglsefteylsesvevps 292

Qy 166 -----AIGGTN-----GS-----FL 175

Db 293 pfdileptsggflklskpcocyifpggrgdsalfavngfnmlnggserkscfwklrlh 352

Qy 176 DSVSQUIDKT---NGLGAMNQPLEVSMGNQDPDKLSTGAKLARDQDPDLLLTRNOQCQPPVA 231

Db 353 drvdslilthigdnlpgnsmldqrkaeleesqsgstnsdwmknlsdpdlgvvflnv 412

Qy 232 TQNTQFFMENQQAQLQMKNLIGFPFGNQQPRMTIRNQQPCL-AMGN----- 277

Db 413 penlknpepn---ikmrslseacftlqylnklsmppeplfrsvngntidpvlfqmgv 468

Qy 278 -QOPWYLIGTPRPA--LVSGNQQLGG-----POGNRPPIFLNHQTCL-----P 317

Db 469 gklemywlnpvksskemqymqwtgtnkkaefilpgqgevdipisyltsvssllvwhp 528

Qy 318 AGNOLYGSPTDMHOLVMSTGGO-OHGLL-----IKN-----OQPGSL-----IRGOQPCVPLI 364

Db 529 a-----npaekiirvifpgnstqynileglekikhldfklqplatqkdlqgvv-tpvv 581

Qy 365 DQQPATPKGFTHLNMQ-----VATSMSSPGLRPHSQSQVP--TTYLHVESVSRILN 413

Db 582 kq-----tklkgradsreslkpaakplpsksvrikskeetpevtkvnhve----- 626

Qy 414 GTTGTCQRSRAPAYDSLOQDTHQGNKYILSHSHEISNGNGCKKALPQNSSLPTPIAKLEE 473

Db 627 -----kppkveskekvmvkkdkpvkte--tkpsvtekevpskee-ppspvkaevak 674

Qy 474 RGSKROYHRANGOTEKHDLNLAQQLAOSQ-----DVERHNSSTCVEYLDAAKTKIKQV 527

Db 675 qatdvkpaakektvkketvkpedkkeekpkveakkedktpikkeekpkkeevkke 734

Qy 528 VQENLHGMPPEVIEDDPTDCARKGKNTASISKASKGNSSPVKTKAEKCIKIVPTTPA 587

Db 735 vkkeikk-----eekpekkevketppkevkkkevkk-eeke-----pkkel 783

Qy 588 KKGRAGRKKSVPVPAHASEIQLMOTPP-----KTPLSRKPKGKGR-KSIQDSGKAR 639

Db 784 kklpkdaksstplseakkpaalkpvkpkkesvkkdsaaagkpkkgkikvikkegkaa 843

Qy 640 GPSGELLQDSIAETIIRYMONLYLGDKEFEQONAWLYKGDGALVPVESKRRPRPRVD 699

Db 844 eaavaavtgatattaavmaaaaglaaipakeleaserslmspedlckdfeelkae----evd 900

Qy 700 IDDEFTRIWNLMGKDEKDEKDKKKEK-----WVEEERRVFERGRADSFIARMHLVQG 755

Db 901 vtkdikpqlleli-----edeeklketpeveayvlgkervtkgpaesdegittteg 952

Qy	756	D-----RRFSPWKSGVDSVLGVFLTQNVDHLSLSSGAFWSLAARPPPKLSSSREDERN	808
Db	953	egeceqtpeelepvekgqvddi-----ekfedagaf-eesstetgdyeeek	996
Qy	809	VRSVVVEDPEG-----CILNLNEIPSMOEK-----VQHPSDMEFVSQVDSGSKEQL	853
Db	997	aeteedeepdeeghehvcsasakhsptedesakaadeayirekresvasgdrraeedm-	1055
Qy	854	RDCSNSGIERNFNLEKSIGNLEEVLSS---QDSFPD-----AIFOSCGRVGSCS	900
Db	1056	---deaiek-gaeqeseeadeedkaedareeeyepekmeadyvmavvdkaaaagae	1110
Qy	901	-----CSKSDAEFTTCERFKTVSGTSQSQTGSPNLSDERICLOGNERPHLY	947
Db	1111	eqvgflttptkqlgadsprepasishdetlpggsesetasdeenre-----dqpeef	1164
Qy	948	EGSG-----DVOKETTNVAQKKPDLEKTMMNKWKSVCFCGQPR	984
Db	1165	tatsgytqstielisseptmdemstprdvmsedetneetspsgef-----	1211
Qy	985	NDTNWOTTPSSSYEOCATRQPHVLDIEDF-----GMQGEGLGYSWMSTISPRVDRYKNK	1037
Db	1212	nltkye---sslysqesykpadvtpingfsegstkdtdgkdnasaastispsspmeedk	1368
Qy	1038	NVPRFRPQSGSVPRFTGQIPTSPHPHELPGMLGSGSSAVQEHQDDTOHQOQDEMKNAS	1097
Db	1269	----fsr--saridaycsevkast-----ldikdisavssekvspskspslspspps	1316
Qy	1098	HLOKTFE-----DLNSSSECLTROSS--TKONTIDGCL-PRODTAEDVVDP-	1141
Db	1317	plektplgersvfnfstlptneiksaeavaevpsvtgvveehcaspedtkle-vvspss	1375
Qy	1142	-----LSNNSSLONILVESNSKNKEQTAVE-----YKETNATILREMKGTL	1182
Db	1376	qsvtgsaghtpyyqsptdekshlptvieiekkppavypvsfedsdakdenetasvpmdenv	1435
Qy	1183	ADGKKP-----TSOWDSL-----RKDVEGNEGROERKNKMWDSIDYEAI	1221
Db	1436	pdespiekvslsprllgesayasaysfisaddkasrgaesfeeksgkqgsdpd---	1491
Qy	1222	RRASISEIAEI-----KERGMNMILAIRIKFDLERIVKHGGIDLEWLRESPPDKAK	1274
Db	1492	--qvapvsemstslsyqdkqegkstddfpikedfggekktdd-----dveamssgpa	1540
Qy	1275	DYLLSINGLGKSVCEVRLLLTUHLN-----APPVDTNV-----G	1308
Db	1541	-lalderklgdvsptqdsvsqfsgfkcdkmsisegtvsdksatpvdgvaedtyshmeg	1599
Qy	1309	RIAVRCMWVPLOPLPESLQJLHLELPVLESYQKFLW-----PRCLKDORTLYELHYQL	1363
Db	1600	vavstasvatssfpe-----pttdvdspslhaevgshstveddalsvsvvgtvp	1649
Qy	1364	ITFGKVCTKSRNCNACPMRGECRHFAVASARLALPAPEERS-LTSAPIPVPESEP	1422
Db	1650	ttfqetempskeecpr-pmsispdfspktaksr--tpvqdhrgsdgsmisiefggcs-p	1705
Qy	1423	PVAIMP-IELPLPLEKSLAGSAPSRENCEPIIEBPSPGQECTEITESDIEDAYYNEDP	1481
Db	1706	eqslamdfrsqsdhptvgagvlhten-----gptevdyspsdmqdslish--	1752
Qy	1482	DEIPTKLNTIQOGMTLREHMERNMELQGDMSKAL-----VALHTTTISITPTPKLNISR	1537
Db	1753	-kipp-----meepsytqndiselsivsgvaeapstssahtp-----sq	1791
Qy	1538	LRYEHQVVELPDSHRLLDGMDKREPDDPSPYLLATWTGTETANSQPPEQCKGKGASKMK	1597
Db	1792	iasplqedtl-----sdvapprmsslya-----	1814
Qy	1598	CFDETSECNLSREANSQTVRGFTLLIP---CRTAMRGSFPLNGTYFQ-----VNELFAD	1648
Db	1815	-----slctekvqsigekjlspskdsislplrpressplytsfdtsavkxekat	1864
Qy	1649	-HESSLKRPIDVPRDWTDWDLPRRTVTFVGTSV--TSIFRLSTEOTO-----FCF	1693

Db	1865	chssspdpdaaasepygf-rasvfdtqmhhlnainrdlstpgleksggktgpdfsay	1923
Qy	1694	WKGFCVCRGFEQKTRAP	1710
Db	1924	qkp-----eetrsp	1933
RESULT 10			
ABGI17148			
XX	ABGI17148	standard; Protein; 2759 AA.	
XX	AC	ABGI17148;	
XX	XX		
XX	DT	18-FEB-2002 (first entry)	
XX	DE	Novel human diagnostic protein #17139.	
XX	XX		
KW	Human;	chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement;	medical imaging; diagnostic; genetic disorder.	
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200175067-A2.		
XX	XX		
PD	11-OCT-2001.		
XX	XX		
PF	30-MAR-2001; 2001WO-US08631.		
XX	XX		
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX	XX		
PA	(HYSE-) HYSEQ INC.		
XX	XX		
PI	Dmanac RT, Liu C, Tang YT;		
XX	XX		
DR	WPI: 2001-639362/73.		
DR	N-PSDB: AAS81335.		
XX	XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity		
XX	XX		
PS	Claim 20; SEQ ID No 47507; 103pp; English.		
XX	XX		
CC	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes.		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. ABG00010-ABG30377 represent novel human		
CC	diagnostic amino acid sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX	XX		
SQ	Sequence	2759 AA;	

Query Match 2.5%; Score 225.5; DB 22; Length 2759;
Best Local Similarity 18.1%; Pred. No. 2.le-07;

Matches 331; Conservative 221; Mismatches 625; Indels 649; Gaps 82;		
Qy	52	VEGPKRKP-----BKPAELPKVVEGPKRPRKAATQEKV-----88
Db	646	lqggsrvptlqgphltnkspasssfqggspsasptvngtqgmgprppqnnplpqg 705
Qy	89	-----KSKETG-SAKKKNLKESATKKPANV 112
Db	706	fqqpyspgnrmvqgnvppnfmvmqgqppnqgppqslhpglgmpkrlpplgfsagqanp 765
Qy	113	GDMNKGPEVTLKSCRA-----LNFLENPGDARGDSESEIVQNSSGANSFSE-----162
Db	766	nfmqgvpsttattcpngsagqlqanqnvhagggagppqng-mqvshppnmmqslm 824
Qy	163	IRDAIGTGNSFLDSVQIDKTNGLGAMNQ-PLEVSMGNQPKLSTGAKLARDQ 216
Db	825	gihnmnnqagtsq-----vpqvnlnsmgqgqgppsqqlmgmhqqlvpsqgqmvqgq- 878
Qy	217	PDLETRNQOCCFPVATQNTOPFMENQOAWLQMKNLIGFP-----FGNQQPRMTIRNOQPC 272
Db	879	-----gtlnpqnmilsraqimpqgqmmvnpssqnlgpsqgmt-----917
Qy	273	LAMGNQOPMYLIGTPRALVSGNQQLGCP-----QGNKRPIFLNH 312
Db	918	-----ppkqmlsgqgqmmaphnqmmgppqgqvlqgnpmieqimtnqmgdnkqfntqn 971
Qy	313	QT-CLPAGNQLYGSPT-DMHOLVMSTGQOQHGLLKKQOP-----GSLIRGQ 358
Db	972	qsnvmpgaqlmrtpnmgmnmvqfvgmsqgmipqgqgvnnpspsqvmglqgvlrppg 1031
Qy	359	PCVPLIDQQATPKGFTHLNQWATSMSSPGLRPHSQVPTTYLHVES-----408
Db	1032	ps-phmaqqhqdpa--ttannndvslsqmmpdvsigqtnmvpv---hvqamqgsasguh 1085
Qy	409	-----SRLTNGTGTQCSRAPAY-----DSLQDDHQ-----N 438
Db	1086	sghmsfnafsgapnqmscgqn--pgfpvkvdxltsppllvnlqgsdisagfhgvnn 1143
Qy	439	KYILSHETSNGCKKALP-----QNSLPTPIMAKLEEARGSROYHRAMQOTEKH 490
Db	1144	k-----qmntnankpkkkpprkkxnsqqlntptdrpagleead-----1183
Qy	491	DLNLAQIAQSDVERHNSCTV-EYDAAKTKIQVQBNLHGMPEVEIEDDPTDG 549
Db	1184	-----qplpgeqglnldnsgkplbefanrppgysqveqrlqgmppqlmqhvapp---1236
Qy	550	ARKGKNTASISKGASKGNSPVKTAKEKCIVPKTAKGKGRKKSVPPPAHASEIQL 609
Db	1237	-----ppqpqq-----pqqlp---qq 1251
Qy	610	WQPTPPKTLRSRKPGRKRSI-----QDSGKARGPSGELLQDSIAEIIYRMQNLVYG 664
Db	1252	qppppsqpsqgqqqqqqqqmmmmmmmqgqpksvrlp-----vsqnvhbpgrplnp 1302
Qy	665	DKEREQONA-----MWLYKDGALVPYESKRRPRKVDIDDETRINWLLMCKGDEKE 719
Db	1303	dsqtrmpmqgsgsvpvmvslgppasvppspdkgrmpmv-----ntplgsnsrkm 1351
Qy	720	GDEKDKKKXWEEERRVFRGRADSFARM-----HLVQGD 756
Db	1352	vygespqp-----ssslpaelnaslpeasgsaepsvpggpnmpshvlpq 1397
Qy	757	RRF---SPWKG-SVVDYSVIGVFLTNVSHLSSS---AFMSLAARFPFKLSSSREDER-N 808
Db	1398	nqlumtgpkpqpslataqatqgppvnslpssghhfpnvaa-ptqtsrpktparas 1455
Qy	809	VRSVVVEDEPCILNLNLEIPSWQEKVQHPSDMEVSGVDSGSKEQLRDCNSGI---ERENF 866
Db	1456	prpyyqtpn-----nrppste-----pseisl-----sperl-nastaglfppqini 1497
Qy	867	LEKSIONLEEVLSQDSFDPALFOSCGRVGS-CSCSKSDAEFPTT-RCETKTVSGTSQS 924
Db	1498	plpprnlrnf--dqglnptttkaigqapsnlmtmopsnfatpqthkldsvvvnsgkqs 1555

Qy	925	VOTGSPNLSDICLOGNERPHLYEGSDVQKQETTNNVAQKKPDLEKTMNMKDSVCFQOPR 984
Db	1556	-----nsgatkraspsnrrssp-----gssr 1577
Qy	985	NDNWKOTTPSSVEQCATROPVHVDIEDFCMGQEGGLGYSHMSISPRVDRVKNKNVPRPF 1044
Db	1578	-----kttsp-----grqn-----skapklitlasqnaa---1 1603
Qy	1045	ROGGSVPREFTGIIIPSTPHELPGMLSGSSSAVQEHODDTQHNQODEMKNASHLQKTF 1104
Db	1604	lqnvlpnrvlvsptplanppvpsfpnnsg-----1634
Qy	1105	DLNNSBECLTROSSYKQNTIDGCLPRDRTAEDVVDPLSNSSL---QNTLVESNNSNKE 1161
Db	1635	-----lnpqnstsvsaaavgvved-----nkeslnvpqdsdcqsqrke 1674
Qy	1162	QTAVEYKETHATILREMKGTLAGKPTTOWDLSLRKDVEGNEGRQ-----ERNKNMDSID 1217
Db	1675	qvnlelkavpa---qevkmvpe-----dqskdggpsdpnkipsveenklvs---1720
Qy	1218	YEAIRRASISEIAIKERGMNMMLAVRIKDFLERIVKDHGGIDLEWLRSPDPKAKDYL 1277
Db	1721	-pamreapt--lsqldnsgapn-----tikppgltdle---vtpv-----1757
Qy	1278	LSIRGLGKSVCEVRLTLHLNLAFFVDTNYGRIAVRMGWVLPLOPLPESLQLHLELYPVL 1337
Db	1758	-vvsgeedkkaasvlp--tlqdlssskpsns-----lnlphsnelcsslvhpel 1803
Qy	1338	ESQKFLWPLRLCKLDQRTLYE-----LHYQITFGVFTKSRPNCNACPMRGECRHA 1391
Db	1804	sevsnnvapslppvmssrpvsststplppnqt---vftv-snp-----1845
Qy	1392	SAYASARLALPAPEERSLTSATIPVPPPEPPVAIPMIELPLPLEKSLASGAPSNRENCE 1451
Db	1846	tsaantaalptlqlsalmstvtvmpnag-----skmvsqggsaagsnar 1891
Qy	1452	PIIEEPASPOQECTETTESDIEDAYNEDDEIPTIKLNIE-----QFGMTLREHMERN 1505
Db	1892	pqfitpvt-----lnssslqvmkgspstipaapltnsglmpbsvavvgplhipqn 1944
Qy	1506	MELQGDMSKALVALHPTTTSITPKLN-----ISRLRTEHOVYELP-----DSHRL 1553
Db	1945	ik-----fssapvpnnalsss-papnigtqrplvissratcpvlpspctssvvpvshpp 1998
Qy	1554	LDGMDKREPDDPSYL-----LAIWTPGETANSQAQPPQKCG-GK 1592
Db	1999	vqvkelnpdeaspqvntsadqntlpsqsstlmvsltnspgssgnrrrpsvsskkgk 2058
Qy	1593	AS--GKMCFTETCSE-CNSLREANSQ 1615
Db	2059	vdkigqlltkackkvtslekggeq 2084

RESULT 11

AAU14697

ID AAU14697 standard; Protein; 5447 AA.

XX AAU14697;

XX AAU14697;

XX 24-OCT-2001 (first entry)

XX Novel bone marrow polypeptide #96.

DE Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;

XX haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;

KW wound healing; nutritional supplement; immune disorder;

KW severe combined immunodeficiency; SCID.

OS Homo sapiens.

XX WO200157187-A2.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PD 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US03782.
XX 03-FEB-2000; 2000US-0496914.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 30-NOV-2000; 2000US-0250683.
XX (HYSE-) HYSEQ INC.
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX WPI; 2001-488875/53.
DR N-PSDB; AAS23002.
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy -
XX Claim 10; Page 124-127; 392pp; English.
XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX Sequence 5447 AA;
SQ
Query Match 2.4%; Score 222.5; DB 22; Length 5447;
Best Local Similarity 17.4%; Pred. No. 1.1e-06;
Matches 322; Conservative 298; Mismatches 701; Indels 527; Gaps 78;
QY 16 TEQNDGSRQDLEFDLNTKTPQKPSKRRKMPKV-----VVEGPKRPRKPAELPKV 69
DB 1594 ttgqdsalqknqsdI-kdiqddignratsfatvkvkdiegfmeenqtklspreital--- 1649
QY 70 VVEGPKRPRKPAELKQKVSKEGSAKKNLKESATKKPANVGDMNSKSEVILKSCRK 129
DB 1650 -----reklhqakeqyealqeetrvaqke-leeavt-----s 1680
QY 130 ALNFDLENPGDARGDSESEIVQNSGANSFSEIRDAIGTNGSFLDSVQIDKTNLIGA 189
DB 1681 alqketekskaak-----elaenkkkidalldwvtsvsgsggqllnlpgmeqlsgasl 1734
QY 190 MNQPLEVSMG-----NQ-PDKLSTGAKLAROQDPDLTRNQOCQPPVATONTQFFPMNOQA 244
DB 1735 ekgaIdtdtgygmvgvnpqekldqcemmkarhqell--sqgnqfilatqsaqafldqgh 1792
QY 245 WL-----OMKNQLIGFFPGNQOPRWTIRNQOPCLAMGNQOPWYLIGTPRALVS----- 293
DB 1793 nltpeeqmIqqkIq-----elkeqysylaqseaelkqvltlqdelqkflqdhk 1842

QY 294 -----GNQOLGGPQGNKRPIFLNHQTCPLPAGNQLYGSPTDMHQLVMSTGQQ 340
DB 1843 efeswlersekelenmhkggsspetlpsllkrq-----gs-----fsedvishkgdl 1889
QY 341 HGLLIKNQO-----PGSLIRGOQPCV---PLIDQOPATPQKFT-----HLNOMVA 382
DB 1890 rfvtisgqvlmdensfkegkepseignlvkdklkdaterytalhskctrlgslhmlilg 1949
QY 383 TSMSSPGLRPHSQSQVPTTYLHVESVSRIILNITGTCORSRAPAYDSLOODIHQGNKYIL 442
DB 1950 -----qyhqfqnseadsIqawmqacean-----veklldstaasdpvvl 1987
QY 443 SHEISNGNGCKKALPONSSLPPTIMAK-----LEEARGSKROYHRAMGOTTEKHDLNLAAQOI 498
DB 1988 qegIattkqlqeelaeh-qypveklqkvardimeiegepdpdhrhvqettdsllshfqsI 2046
QY 499 AQSDQVERHNSCTVEYLDAAKTKIOKV-----VOENLHGMPPEVIEIEDDPTDGARK 552
DB 2047 s-----yslaerssllqkaIaqsqsvqdslesllqslsigeveqn-----le 2086
QY 553 GKNTASISKASKGNSSPVKKTAEKKEKCIIVPKTPAKKGRAGRKKSVPVPPAHASIQIWLQWP 612
DB 2087 gkvssIsgvi-----qeatnmklkqdiarqkssl--eatremvtfme 2131
QY 613 TPKTPTLSRSPKPGKGRKSIQDSKARGPSGELLQCDQSIABIIYRMQNLVLDGKEREQEQ 672
DB 2132 tadsttaa-----vlgqlaevsqrfegqlclqqek----- 2162
QY 673 NAMVLYKGDGALVPYESKKRPPKVIDIDETT-RINWLLMGKDEKEGEDEKDKKKEK 731
DB 2163 -----essllklpqaemfehlqgklqgfmenksrmlasngnqpdqdlthf 2207
QY 732 WEERRRVRGRADSFIARMHLVQGRDRFPWKGVSDVSVIGVFLTONVSDHLSSAFMSL 791
DB 2208 fqq-----iqelnlemedqgen-----ldtl-----ehlvtelsscgfald 2243
QY 792 AARPPKLSSSRREDNRVSVVVE-----DPEGCILNLNE-----IPSWOEKVO---HPS 838
DB 2244 lcqhqrdivgnlrkdftelqktvkeredasscqqldefklvrtfkqkwiketesippt 2303
QY 839 DMEVSGVD-----SGSKEQLRDCNSG--TERFNFLEKSIQNLSEEVLLSSQD-----SFD 886
DB 2304 etmsakelekqtehlkllldwaskgtlveelnlykgtslenlmeitapdsqgktgsil 2363
QY 887 PAIFQSCGRVG---SC-----SCSKSDAEFP-----TTRCET---KT 917
DB 2364 psvgssvgsvngyhtckdlIteiqcdmsdvnlkyeklgglvherqesIqalInrmeevhke 2423
QY 918 VSGTSQSVQT-----GSPNLSEICLOGNER-----PHLYEGSGDVQ---KQETT 960
DB 2424 ansvlqlweskeevlksmdamssptkktetvkaqaesnkafIaeleqpskIqkvkealag 2483
QY 961 VAQKKPDLKTMNWKDSVCFQPRNDTNW-----QTTPTSSSYEQC-----ATR 1003
DB 2484 llvtypnsqaeenwkkl-----qeelnrweratevtvarqrleesashIacfqaaesql 2539
QY 1004 QPHVLDTE-DFGMQGEGLGYSWMSISPRVDRVKNKNVPRFRFGQSGVPREFTOGIIPST 1062
DB 2540 qpwlmeIkelmmgvIgp-----Isidpmlnaqkqv--qfmIkefearrqheqlneaa 2591
QY 1063 PHELPGMG-LSGSSSAVQEHQDDTQHNOQDMMKNKASHLQKTFDLDLNSSEECITRQSSSTK 1121
DB 2592 qgiltgpgdvslstsqvqke-----lqslnqkwelldklnsr 2629
QY 1122 QNITDGLCPDRDRAEDVVDPLSNSSL--QNILVESNSSKEQTAVYKTNATILREMK 1179
DB 2630 ssqIdqalvktstgyqellldsekvragvqrIsgsaIstqpeavkqgleetseIrsld 2689
QY 1180 GTLADGKKPTSQWDSLRKDYEGNEGROERNKNNMDSIDYEAIRRASISEISEAIKRGMN 1239
DB 2690 qldhevkeaqIcIdelsvll-geqylkdelkrlIetv-----aIplqgle 2733

QY 1240 NMLAVRIKDFLERIVKHGIDL-----EWLRESPDPKADKYLLSIRGLGKS-----VE 1289
Db 2734 dilaadrinriqaalastqfqcmfdeirrtwldkqsaqncpisaaklerlqslqene 2793
QY 1290 CVRLTLUHLNAPVDNNGRIANVMGWVQPLPESLOLHLELIPVLESIQFLWPRLC 1349
Db 2794 fqalsnhdsgsyvivaegesllils--vppgeekrtlqnlvelknhweelskktadrqs 2851
QY 1350 KL----DQRTLYELHYOLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPE 1405
Db 2852 rlkdmqkaqkyqhve-----dlvpwiedck---akmselrtvld-- 2889
QY 1406 ERSLSATIPVPPSPFPVPAIMIEPLPLEKS--LASGAPSN-----RENCEPIIEEPASP 1460
Db 2890 -----pvqllessllrskamlnevekrslleilnsaadi 2923
QY 1461 QGECTETESDIED--AYYNEDDP-----EPTIKLNIEQFGMTLREHME--RNWELQ-EG 1511
Db 2924 linsseadedgirdkaginqnmavteeelqaktgsleemtqrlrefqesfkniekveg 2983
QY 1512 DMSKALV--ALHPTTTSIPTPKLKNISRLRTEHOVYELPDS-----HRLLDGMDKREPD- 1563
Db 2984 akhgleifdalgsqacs-----knleklraqgevlqaleqgvdylnftqgivedapdg 3038
QY 1564 -DPSFYLLAIWTPGETANSAPPEQKCGKASGKCMCFDETCSECSLREANSOTVRGTL 1622
Db 3039 sdasqll-----hqaevaagqe-----flevkqrwns-----gcvm 3068
QY 1623 IPCRTMRGSPFLNGTFOYNELFADHESLSKPIDVPRDWTWDLPRPT 1670
Db 3069 menkleglqgf-----hcrvremf-----sqliadideldmggaigrdt 3107

RESULT 12
ABB67502
ID ABB67502 standard; Protein; 3257 AA.
XX ABB67502;
AC
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 29298.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL11605.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 29298; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3257 AA;

Query Match 2.4%; Score 222; DB 22; Length 3257;
Best Local Similarity 17.6%; Pred. No. 5.2e-07;
Matches 349; Conservative 273; Mismatches 775; Indels 586; Gaps 77;

QY 7 SSAVNATEATEQNDGSRQDVLFEFLNKTPOQKPSKRKRKF-----MPKVVEE 53
Db 798 ssskdeppaeenlpapdqdpie--qgktpvakngqhdkehneapkaeslsvsdipssvt 855
QY 54 GKPKRPRKPAELPK--VVVEGKPRKPKAATQEVKSKETSGAKKKNLKESATKK-- 108
Db 856 pskrnhsspantpkkskeialealqssvrralrsdk-----atpqniresrskrtl 907
QY 109 -----PANVGDMSNKP---EVLTKSCRKAL-----NFDLEN 137
Db 908 teitllmddtmrrssprlgrspaeshsherspmekktvtvsklakdlitidkekeitelks 967
QY 138 PGARQGD-----SESEIVQNSSGANSFSEIRDAIG-GTNGSFLDSVSQIDKTNG 186
Db 968 lpadsetkdvkikttaasdtltdenpssttemkklgkplkakkmartetevkka 1027
QY 187 LGAMNQ--PLEVSMGNQPKLSTGAKLARDQOQDLTLRNOQ--COFPVATQTFPMENQ 242
Db 1028 iadsnedipsifskvceehltssesegkdekeellcpkpqidctndtldqstaletde 1087
QY 243 QAWLQMKNLIGFPFGNQ-----OPRMTIRNOQ--PCL 273
Db 1088 qveekrsnrirsrirnekkftetdtlsdhldakkaenasleismrpkctietqgsqpt 1147
QY 274 AMG-----NQOPMYLIGTPR-----PALVSGNQQLGGPQGNKRPILFNH-----QT 314
Db 1148 akknrsgrlrsrkeksvinaaksekdkpsais-----qsterqllnepsdkdkk 1199
QY 315 CLPAGN--QLYGSPTDMHQLVMSTGGQOHLIKNOQPGSLIRGQQPCVPLIDQOPATPK 372
Db 1200 teqsgnkkkavvgpldktetsstn-----iiddknesfidsamqpsdrlnqkesa-- 1250
QY 373 GFTHLNQWATSM-----SSPGLRPHSQSQVPTTYLVHVES----- 407
Db 1251 -ftklssissppkkimkdqdkldalsksgdsnptirdtgedsrqtdkdkhgendckheeed 1309
QY 408 ---VSRILNGTTGTCORSRAPAYVDSLQODI-----HOGNKVILSHEI 446
Db 1310 ssklikanidetkssekdaepiskdssqdsakprlskpknrknkkkpkndksiaesdi 1369
QY 447 SNG-----NGCKKALPQNSSLPTPIMAKLEEARSGKR--QYHRAMGQT 487
Db 1370 eggfqvtetvqatcstpsesnkkdmvksdetneepnlseteigrirkrqgafhiempkd 1429
QY 488 EKH-----DLNLAQIIAQSQDVERHNSSTCV-----EYLDAAK-KTKI----- 524
Db 1430 dlhitpqnqnsiagvnfekqvplpesve---sdtplmktpktylmctknktsllsase 1486
QY 525 -----QKVQVQNLHGMPEVIEIDDDPTDGARKGKNTASISKGASKGNSSPVKKTAE 576
Db 1487 dpdivleqqlittsksgdsnpdlndannletsstqdpke-hefsdqftdndsdiipsc 1545
QY 577 KEKCIIVPKTPAKKGRAGRKKSVPV-----PAHASEI 607
Db 1546 ksqivftpttkssdqtknsfitnrsppkskrwnsveakrldnsfeesqnaasessaskv 1605
QY 608 QLWQFTP-----PKTPLSRSPKKGKRSIQDSGKARGPSGELLQODSIAEII 656

Db 1506 qkelrtptascrkrlvrlktrpts-slpt-nsrksifktpaksckrltkil-esmektps 1662
Qy 657 RMQNLVYG---DKEREQEQNAMVLYRGDQALVPYESKRPKPKVDIDDDETRWNLLM 712
Db 1663 repvsIgevnpsdvpaaesavlhesdrdlesnel-----pneevfedt----- 1708
Qy 713 GKDEKEDGDEKDKKKKWEERERFRGRADSFIAHMLVQGDRRFPSPWKSGVSDSVIG 772
Db 1709 ---eaaaeatdnklkkeddhellev-----ndicaasknpltdtdskdassnksdtdsv- 1760
Qy 773 VFLQNVSDHLSSAFMSLAARFPKPLSSREDERNVRSVVVDEPEGCILNLNEIPSWQE 832
Db 1761 ---lqetkdelnslnatqgdtpikelteevevpnnkt-ved----- 1799
Qy 833 KVQHPDMEVSGVSGSKEQLQDCNSGIERFNFLEKSIQNLKEEVLVS---SODSFDPAI 889
Db 1800 -----eskkqellkd-----lepdnaaleedtaastakaaeamdyi 1835
Qy 890 FQSGCRVSGSCSKSDAEFTTTRCETKTVSGTSQSVOTGSPNLSDEICLOQNERP----- 944
Db 1836 ke-----ksnvksvlaepet-dvtddeelaqspipnssettsvtdtdpepstsv 1883
Qy 945 ---HLYEGSDVQKQETTNVAOKKPDLEKTMWWDKSVCFQGPNDTWNQWTPSSSYEQCA 1001
Db 1884 vksrlrkreadssqpdeaaakrkqrdveksltgk-----qvkpa----- 1924
Qy 1002 TROPHVLIDIEFGMQEGGLGYMWSISPRVDVKNKNVPRFRPQ-----GGSVPREF 1055
Db 1925 -rrqlaeeve-----rpslkrkseteakstvggkyisignetsms 1968
Qy 1056 GOIIPSTPHELPGMLSGSSSAVOE--HODTOHNOQDEMKN-----ASHLQKTFLD 1105
Db 1969 tapiretnreaastpsarksavqeahvethkhiilpgpgkllhdsdpaaevkqpmvq 2028
Qy 1106 LNSSECLTRQS-----STKQNTDCLPRDRTAEDVDVPLSNSSL----- 1148
Db 2029 tllsstlsqkpsltdgspklrkstlksksladienldgdsfsssvlnkntsvapr 2088
Qy 1149 QNILVE--SNSNKEQAVEYKETNATILREMKGTADGKKPTSQWDSLRDKRDEGNGRQ 1206
Db 2089 vnvisllqsktdqvetaasasetpiltkktktstkkp-----egnkkte 2137
Qy 1207 ERNKNNDSDY-----EATRASIASEISAIEKERGMNNMLAVRIKDFLERIVKDHGGI 1260
Db 2138 skkkslvqpgmqktqkeeeavsgpki--lnkylksetessrktvtstgrkqig----- 2190
Qy 1261 DLEWLRSPDPKADYLLSIRGLKSVCEVRLTLNLNLAFFPVDVTNVRTAVRMGWVPLQ 1320
Db 2191 -levlkkpesrkseesi--vealsrkkqsgvqriskidgrkseqtslpqpdvsksetalk 2247
Qy 1321 -PLPESLQLHLELYPPVLESTQKFLWPLRCKLDQRTLVLYQLITTFKGVCFKSRPNCN 1379
Db 2248 aalpkete-----fpvqda-----eiekmskgrghqnavkntkteqpskpkte 2291
Qy 1380 ACPMRGECR-----HFASAYASARIALPAPERS 1408
Db 2292 vrsIqaeaatelmdsmdsqsdvsdiratfpesqgifnvgpghmtraissnrsIaptptms 2351
Qy 1409 LTSATIPVPPSPFPVAIPMTLEPLPLEKSLASGAPNRENCEPIIEEPASPGQECTEI- 1467
Db 2352 dsqrn--askerftcpvsdq--kkipresqtlskrarrgrnqnpivskrkageaedgtavi 2407
Qy 1468 -----TESDIEDAYNED-----PDEIP-----T 1486
Db 2408 npkprremdeedhpqgdhqvqesaafaafpvkitaassvipqvvrstgntvpqnisprklc 2467
Qy 1487 IKLNIEQFGMTLREHMRNMELQEGD---MSKALVALHPTTTSIPTPLKNI---SRLRTE 1541
Db 2468 vklnrrpynkwlrtqernee-qegsnvtslpllgsetdsaaesmsesilqsgvqse 2526
Qy 1542 HQVYELPDSDHR---LLDGMOK--REDDPSPYLAIWTPGETANSAOP-----PEQKCGG 1591

Db 2527 paiglpasqsdctlqasdlrriressaqIapiaaaydsaaandstspaldiapesaqta 2586
Qy 1592 KAS 1594
Db 2587 kat 2589
RESULT 13
ID ABB65772 standard; Protein; 5533 AA.
XX ABB65772;
AC ABB65772;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 24108.
DE Drosophila melanogaster polypeptide SEQ ID NO 24108.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL09875.
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.
PT Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 5533 AA;
SQ
Query Match 2.4%; Score 221.5; DB 22; Length 5533;
Best Local Similarity 19.1%; Pred. No. 1.4e-06;
Matches 352; Conservative 245; Mismatches 695; Indels 551; Gaps 83;
Qy 5 MDSSAVNATATEQNDGS-----RQDVLFEFLNKTPOQKPSKRKKFMPKVYVE 53
Db 1362 insslnasgmgsccstflpspsrywssshhqnnhqgqqlhgssssntclm 1421
Qy 54 GKPKRKR-----KPAELPKVVVEGPKRKRKAATQEKVKSKETGSAKKKLNKESATK 107
Db 1422 aspar-prslssnsdsdvpqnaaggslderlrntfeenywsggsre----- 1471
Qy 108 KPAVNGMSNKSPEVTLLKSCRKALNFLENPGDARQGDSESEIVQNSGSFSEIRDAI 167

Db 1472 -----hishtpssatpwsqglmhmlnltglnshqtssagnsnssgt-----vsssa 1520
Qy 168 GGTNGSFLDSVQIDKTNGLGAMNQPLEVSMGNQPKLSTGAKLARQQPDLLTRNOCCQ 227
Db 1521 snsrhkfld-----idel-----qpsdiv-----ksvialkvsfddqfrlnkqwyd 1563
Qy 228 FPVATQNTQPMENQOAWLQKNOLOGIFPF-----GNOQPRWTIRNOQPCLA 274
Db 1564 -----psssdalgsnssnltvgsslvavsrhpggpcsgntspal-----pula 1607
Qy 275 MGNOQPMYLGTTPRALVSGNOQLGGPQGNKRPTFLNHQTCPLPAGNOLYCSPTDMHQLVM 334
Db 1608 atkatp-iilgn-----csdg--lgnstgskagilqlrlssl-----spmspqasm 1650
Qy 335 STGGOQHLLIKNOPCSLRGQPCVPLIDQOPATEPKGFTHLNOMVATSMSSPGLRPHS 394
Db 1651 sp-----ynspspspvgvgtacilqitkpaaptasagisggtaasssp-----a 1697
Qy 395 QSOVPTTLHVESVSRILNQTTCQSRAPAYDSLOQDIHQGNKYITLSHEISNGNCKK 454
Db 1698 ansqtkglgyfpshpplntaappavpappplpe---mgkqsrltgg-sggnlntk 1753
Qy 455 AL-----PONS-----SLP-----TPMAKLEEARGSKROYHRAMGOTEKHDLN 493
Db 1754 slsvdpdqpqsparvqlqksasvpgstnvaapslsldstasvetsasissstsgnss 1813
Qy 494 L-----AQIAQSQDVERHN---SSTCVBYLDAAKTKIKQVVOENLHGM-----PPE 538
Db 1814 ltsaaihvpqpgstfveehtksgtstsgssssskkiss-thdklshkhnrrseadk 1872
Qy 539 VIEIEDPTDGARKGN--TASTISKAS---KGNSSPVKTAKEKICVPTPAKKGRAG 593
Db 1873 kikkdknaassdkrknstsgskstaprieddsaadtdak-----aekng-- 1921
Qy 594 RKKSVPAPHAASEIQLQOPPPKTPLSRKPCKGKRSIODSGKARGPSELLCOODSIAE 653
Db 1922 -----rhekekerqekrekdlrkqvereekdrkaqgeerek-----edrkak 1964
Qy 654 IIRYMQNLYLGDKEQEQONAMVLYKGDGALVPYVESKKRKPVRKVDIDDETRIWNLIMG 713
Db 1965 -----eeekererekaq-----edrekkereerelekeqr-----d 1997
Qy 714 KGDEKGEDEKDKKKKEWEEERVRFRGRADSFTRMHLVOGDRRFPKGVVDSVIGV 773
Db 1998 keqkeleirekdlrekeqrdrnrekelr-dkdlrekekrekeqr-----ekelhr 2047
Qy 774 FLTQNVSDHLSSAFMSIAARFPKLSRSSREDERNVSVVDEDEGCILNLNETPSMOEK 833
Db 2048 ekdqrereh-----rekeqsrmdvqeg-egrggrmrelssyqk- 2095
Qy 834 VQHPSDMEVSGVDSG-----SKEQLRD-----CSNSGIERFNFLEK-S 870
Db 2086 ----skndiagaeasltaidcqhnkenamtdiagtgpaapstpsdntpkersklrns 2141
Qy 871 IONLEEVLLSQDSFDPALFQSGRGVSCSKSDAEFPPTRCETKTVSGTSQSVQGTSP 930
Db 2142 pvrlnhrrlssqes-----nhsaggsgcgshqhiedyvkfirmensqnishvsnq 2196
Qy 931 NLSDEICLOGNE-----RPHLYEGS-----GDVQKQET 958
Db 2197 rlndrrskshkssfkedknssshsrphgsggssassskhhrrdkhhqkgssastet 2256
Qy 959 TN-----VAQKKPDLEKTMNKKDSVCFQGNRNDTNWQTPSSSYEQCATRQP-HVLD 1009
Db 2257 nssievvvdpisqtkhnl-----ntseeelqshqpkreke 2291
Qy 1010 IEDFGMOGEGLYGSMWISPRVKNVKNVPRFRFGGVSVPREFTGOIIPSTPHELPGM 1069
Db 2292 rehtshanssssrhks---krdhhhrrekhrhsvaestntdehtbpq--qhnphrrisa 2346
Qy 1070 GLSGSSSAVQEHQDDT-----QHNOQDENMKASH-----LQKTFLLDLSSE 1111
Db 2347 agsgsagelssaatntsgklhhdhrrsverkssrgsgdeghhsskslraklmmllssad 2406

Qy 1112 ECTROSSTKQNTDGLCLPRDRTAEDVVDPLSNSSNQILVSNSSNKEQTAVEYKETN 1171
Db 2407 sdtddaskkhsifd--ipddcpnvsmysk-----vkarscknmqraeeekik 2453
Qy 1172 ATILREMKGTLAGCKRPTS-QWDSLRKDVEGNEGRQERNK-----NNMDSIDY 1218
Db 2454 ak-fsqliksrakkrstsydgs--dte-fedqrhnsghsssfhgyrpglssssddddd 2508
Qy 1219 EATRASSISISEAIEKRGMMNLAV-----RIKOFLEIRIVKHGGIDLEWLRSPDK 1272
Db 2509 eethgrissdsda-ehggdngdgastladanvrqmqqlrlldcgdd-----ssede 2561
Qy 1273 AKDYLLSIRGLKSGVECVRLTLHLNLAFFVDTVNGRIAVRMGVPLQPLPESLQLHLE 1332
Db 2562 irrnmvkhshfgkrnsnstrias-----dse-----sqsqpapd--lciqg 2600
Qy 1333 LYPVLESIQFLWPRICKLDQRTLYELHYLIYFGKVFCTKSRPNCNAC-----P 1382
Db 2601 ehplapa-qeikreqisdeeq-----fksrhdnsnsseerklktere 2642
Qy 1383 MRGECRHF--ASAYA-SARLALPAPEERSLTSATIPVPPEFPPVAIPMIELPLPLEKSL 1439
Db 2643 iktelgdgfynsseytytgkikespetrkkhks-----krrl 2680
Qy 1440 ASGAPSRENCE-PIIEEPASP-----GOECTETTESIEDAYNEDDEPTTKLNI 1491
Db 2681 kssstadtsaaqplvmtpltpstpsifdvhssseck--tkfd----nfdlktecsslplel 2734
Qy 1492 EORGMTLREHMERWERNMELQE--GDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPD 1549
Db 2735 s--agerikrkkerekkreklrmteatvpnsptndtsetkskeer----- 2780
Qy 1550 SHRLDGMKREPDD-----PSPYLLAIWTPGETANSAQPPEQKCGKAS 1594
Db 2781 -hrllkskkskmdnsctkiynssgahpstpslpa--tptsapstaqlsk-----rge 2832
Qy 1595 KMCWF-----DETSCNSLREANSQTVRGTLIPORTAMRG 1631
Db 2833 dknefifgiisdeesqfpegaetnkq-----ilpsvsttg 2869
RESULT 14
ABBT1160
ID ABB71160 standard; Protein; 5560 AA.
XX ABB71160;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 40272.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR N-PSDB; ABL15263.
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

PS Disclosure; SEQ ID NO 40272; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU01840-ABU01840-ABU01840), expressed DNA
CC sequences (ABU01840-ABU01840-ABU01840) and the encoded proteins
CC (ABU01840-ABU01840-ABU01840).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5560 AA;

Query_Match 2.4%; Score 221.5; DB 22; Length 5560;
Best Local Similarity 19.1%; Pred No. 1.4e-06;
Matches 352; Conservative 245; Mismatches 695; Indels 551; Gaps 83;

QY 5 MDSSAVNATEATEQNDGS-----RQVLEFDLNTKTPQOKPKRKRKFKMPKVYVE 53

Db 1362 lnnsslnasqmgscgstflpspsrywssshhqnqnnhqgsgqlghssssntclm 1421

QY 54 GPKPKRPR-----KPAELPKVVVGKPKRPRKATQEKVSKETGSKAKKNLKEATK 107

Db 1422 aspar-prlsnsdsdsvpgnaggsplderlfnfeenywrgssre----- 1471

QY 108 KPAVGDMSNKSPEVTLKSCRKALFNLENPGDARQDSEISIVQSSGANSFSEIRDAI 167

Db 1472 -----hisghpssatpawqlsmhmmlstglnshqtasagssnsstg-----vssa 1520

QY 168 GGTNSFDSDVSDIKTNGCLGAMNPLEVSMGNPDKLGSTGAKLARDQDPDLLTRNQCC 227

Db 1521 snsrhkfid-----idel-----qpsdiv-----ksvlakksvfdldfgrlnknqwyd 1563

QY 228 PPVATQNTQFPMENOQAWLQMKNLQIGFPF-----GNOQPRMTIRNQPCLA 274

Db 1564 -----pessafalgsssnivtgselvanvahrpgpcsgntspal-----pala 1607

QY 275 MGNOQPMYLIGTPRALVSGNQQLGGPOGNKRPFLNHTOTCLPAGNQLYSGPTDMHQLVM 334

Db 1608 atkatp--lign-----csgg--lgnstgskagllqlrsl-----spmsppqasm 1650

QY 335 STGGQGHLLIKNQOPGSLIRQOPCPVLIQOPATPCKGFTHLNQMVATSMSSPGLRPHS 394

Db 1651 sp-----ynspspsvgvgtacglgtkpaapgtasaglggtasassp-----a 1697

QY 395 OSOVPTTLHVESVSRILNGTGTQCRAPAYDSLOQDIHQGNKYILSHEISNGNGCKK 454

Db 1698 ansqgtklqyfpshpplntcaappavqappipe-----mgksritgq--ssgnnltk 1753

QY 455 AL-----PQNS-----SLP-----TPIMAKLEEARGSKRQYHRAMQTEKHDLIN 493

Db 1754 slsvdpqqssparvqlqksavpgstnvapsslsldsttasvetsasistangns 1813

QY 494 L-----AQQIAQSDVERHN---SSTVEYLDAAKTKYIKVQENLHGM-----PPE 538

Db 1814 ltsaiahvqkpgqgtfveehcktsqstgsssssskiss-thdklshkhnrsesdk 1872

QY 539 VTEIEDDPTDGARKGN--TASISGAS--KGNSSPVKKTKAEKECIVPTKAKKGRAG 593

Db 1873 kkkdsdnassdkrknstsgssksatpridedsseaddtadk-----aekng-- 1921

QY 594 RKKSVPFPAHASEIQLWQPTTPKPLSRKPKGKRKSIQDSGKARGPSGELLCCODSTAE 653

Db 1922 -----rhekekerqekredlirkvreeekdrkaqqearek-----edrtrak 1964

QY 654 IIRYONLYGDKEREQEQONAMVLYKGDGALVPYESKKRPKPKVDIDDETTRINWLLMG 713

Db 1965 -----eeekerrekkaq-----edrekererelekeqr-----d 1997

QY 714 KDEGEDEKDKKKKEKWEERVRFRGRADFIARMHLVQGDGRFSPWKSGSVSVGV 773

Db 1998 keqkekeirekldrekegrdrnrekeir-dkdlrekenrekeqr-----ekelhr 2047

QY 774 FLTONVSDHLSSAFMSLAARPPKLSRRDERNVRVSVVVEDPEGCILNLEIFSWQPK 833

Db 2048 ekdqrereh-----rekegrramdveq-egrgmrrelssyqk- 2085

QY 834 VOHPSDMEVSGVDGSG-----SKEQLRD-----CSNSGIERFNFLEK--S 870

Db 2086 ----skmdiaageassltaldcqhnkenamdtiaqgtgaspstpsdntpkersrklsrns 2141

QY 871 IONLEEVLSQDSFDPALFQSCGRVGCSCSKSDAEFTTRCETKTVTSGTQSQVOTGSP 930

Db 2142 pvrllhkrllssqes-----nhsaaggsgcgssghqihedyvkrirmensqaisvhsnq 2196

QY 931 NLSDEICLQGN-----RPHLYEGS-----GDVQKQBT 958

Db 2197 rlndrrdskehsksfkedknssshisrphgcgssasasskhhrrdkhhqkgsassiet 2256

QY 959 TN-----VAQKPDLEKTMNWKDSVCFQOPRNDTNMQTTTPSSSYEQCATROP-HVLD 1009

Db 2257 nssievvvdpisqtkhnl-----ntseeelqshqkpreke 2291

QY 1010 IEDFQMGQEGCLYSWMSISPRVDRVKNKVPFRFRQGSVPREFTGIIIPSTPHELPGM 1069

Db 2292 rehfsshanssrrhks---krdhhhrkkrhsvaesntdehtpg-qhaphrrisa 2346

QY 1070 GLSGSSSAVOEHQDQT-----QHNOODEMNKASH-----LQKTFDLDLNSEE 1111

Db 2347 aqsgsagelssaataintsgklhqqhrrsverksgsdegghsskskslraklmlmsaad 2406

QY 1112 ECLTRQSTKONITDGLPRDRTAEDVVDPLSNSSSLONILVESNSSNKEQTAPEYKETIN 1171

Db 2407 sddtdaaskkhsifd-ipddcpnvsmlyd-----vkaracknmrgqaeekklk 2453

QY 1172 ATILREMGTLADGKKPTS-QWDSLRLKQVNEGEGQERNK-----NNWDSIDY 1218

Db 2454 ak-fsqllqsktrakkrrstsydgds---dte-fedqhrnsgsssfhrypglsssdodd 2508

QY 1219 EAIRRASISEISEAIKERGMNNMLAV-----RIKDFERIVKDHGIDLEWLSPDPK 1272

Db 2509 eethrriissdsda-ehaggqdnqgastladanrvqmqlrllcdgdd-----ssede 2561

QY 1273 AKDYLLSIRGLKSVCEVRLTLNLNLAFFVDNTNVRGRIAVRMGWVPLQPLPSQLHLE 1332

Db 2562 lrnmvkhshfgrksnsstrias-----dse-----sgsqgpad---ltikq 2600

QY 1333 LYPVLESTOKFLWPRLCKLDQRTLYELHYQLITFGKVFCTKSRPNCAC-----P 1382

Db 2601 ehpiapa-geikreglsdeeq-----fksrhdnsnsieerklktere 2642

QY 1383 MRGEGRHF--ASAYA-SARLALPAPEERSLTSATIPVPPSPPPVAIPMIELPLEKSL 1439

Db 2643 ktelgdgfymsseytytgkkeysptekkkks-----krrl 2680

QY 1440 ASGAPSNRENCEP-PIIEBPASP-----GOECTEITSDIEDAYYNEPDPDPIPTIKNI 1491

Db 2681 kssstadsaaqtpivmtpltpsfidvhsseck--tkfd-----nfdldlkeccsplei 2734

QY 1492 EQFGMTLREHMERNELOE--GDMSKALVALHPITTSIPTPKLKNISRLRTEHQVYELPD 1549

Db 2735 s--agerrkkrkkrkkrlnmtcatvpsnptndtsseklskeer----- 2780

QY 1550 SHRLDGMKRPDD-----PSPYLLAIWTGCTANSQAOPPKCGGKAS 1594

Db 2781 -hrkkkskksmdnsontklynssgahpstpsipa--tptsapstaqtsk-----rge 2832

QY 1595 GKMF-----DETSECNSLREANSQTVRGTLTLLIPCTAMRG 1631

Db 2833 dkmeffilgildeesqfpeaetnkd-----lipssvsttg 2869

RESULT 15

AAU14603
ID AAU14603 standard; Protein: 5373 AA.

XX AAU14603;

DT 24-OCT-2001 (first entry)

XX Novel bone marrow polypeptide #2.

XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
XX haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
XX wound healing; nutritional supplement; immune disorder;
XX severe combined immunodeficiency; SCID.

OS Homo sapiens.

XX WO200157187-A2.

PN 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03782.

XX 03-FEB-2000; 2000US-0496914.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 30-NOV-2000; 2000US-0250683.

XX (HYSE-) HYSEQ INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;

PI Ren F, Drmanac RT;

XX WPI: 2001-488875/53.

DR N-PSDB: AAS22908.

XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic

PT and gene therapy -

PT Claim 10; Page 198-208; 392pp; English.

PS AAU14602-AAU14794 represent novel bone marrow polypeptides of the

XX invention. The proteins and corresponding coding sequences may be used

CC in the prevention, diagnosis and treatment of diseases associated with

CC inappropriate bone marrow polypeptide expression. For example, to treat

CC disorders associated with decreased expression by rectifying mutations

CC or deletions in a patient's genome that affect the activity of the

CC polypeptides by expressing inactive proteins or to supplement the

CC patient's own production of the polypeptide. Additionally, the nucleic

CC acids may be used to produce the polypeptides, by inserting the nucleic

CC acids into a host cell and culturing the cell to express the protein.

CC The nucleic acid and its complementary sequences may also be used as DNA

CC probes in diagnostic assays to detect and quantitate the presence of

CC similar nucleic acid sequences in samples, and therefore which patients

CC may be in need of restorative therapy. The proteins may also be used as

Query Match 2.4%; Score 216; DB 22; Length 5373;

Best Local Similarity 17.5%; Pred. No. 3.6e-06;

Matches 322; Conservative 292; Mismatches 698; Indels 524; Gaps 77;

Qy 16 TEQNDGSRQDVLEFDLNKTPQOKPSKRKRKFKPKV-----VVEGKPKRKRKPAELPKV 69
Db 1542 ttqddlsalqknqsdll-kdlqddiqnratstfatvkvkiegfeenqtklspreltal--- 1597
Qy 70 VVEGKPKRKRKKAATQBKVKSKETGSAAKKNNLKESATKKPANVGDMGNKSPSEVTLKSCRK 129
Db 1598 -----reklhqakeqyealqeetrvaqke-leeavt-----s 1628
Qy 130 ALNFDLENPGDARQDSESEIVQNSGANSFSEIRDAIGTNGSFLDSVSOIDKTNGLGA 189
Db 1629 alqgeteksaak-----elaenkkidalldwvtvsgsggqllnlpgmeglsasl 1682
Qy 190 MNOPLEYSMG---NO-PDKLSTGAKLARQDQPDLLTRNOQCQFPVATQNTQFFMENQQA 244
Db 1683 ekgaltdtdgymgvnqapeklldkqcemmkarhqell--sqgnfilatqsaqafldqghg 1740
Qy 245 WL-----QMKNLICFPFGNQOQPRMTIRNOQPCIAMNQOQPMYLIQTTPRALVS----- 293
Db 1741 nltpeeqmllqklg-----elkeqystslagseaealkqvltqdelqklqdhk 1790
Qy 294 -----GNQQLGGPQGNKRPIFLNHQTCLPAGNQLYGSPTDMHQLVMSTGGQQ 340
Db 1791 efeswlersekelennhkggsspetlpsllkrq-----gs-----fsedvishkgdl 1837
Qy 341 HGLLIKNOQ---PGSLIRGOQPCV---PLIDQOQPATPKGFT-----HLNOMVA 382
Db 1838 rfvtsigqkvldmensfkegkpeignlvkdkldaterytalhsketrllgshlnmlig 1897
Qy 383 TSMSSPGLRPHRSQSQVPTTYLHVESVSRILNGTTGTCQSRAPAYDSIQDHOQNKYIL 442
Db 1898 -----qyhqfqnssadsqawmqacean-----veklstsvsdpgvl 1935
Qy 443 SHEISNGNGCKKALPONSSLPPTIMAK-----LEEARSKROYHRAMGOTEXHDLNLAQOI 498
Db 1936 qeqlattkqlqeelaeh-qvpveklqkvardimelegepdpdhvqgettsilshfql 1994
Qy 499 AOSQDVERHNSSTCEVYLDAAKTKIOKV-----VOBNLHGMPEVEIEIDDDTDGARK 552
Db 1995 s-----yslaerssllkkaiaqsqsqeslesllqsigveqpn-----le 2034
Qy 553 GKNTASTISKGASKGNSPVKKTAEKECIVPKTPAKKGRAGRKSKSVPPPAHASEIQLNQOP 612
Db 2035 gkvyslssgvi-----qkalatnmklkqdiarkksal--eatremvtrfme 2079
Qy 613 TPPKTPLSRSRPKGKGRKSQDSGKARGPSGELLCDQSIABLIYRMQNLILGDKREREQO 672
Db 2080 tadattaa-----vlgklaevsqrfeqlclqgqek----- 2110
Qy 673 NAMVLYKGDGALVPVEKRRKRPKVDIDDETT-RIMNLLMGKDEKGEDEKDKKKEKW 731
Db 2111 -----esllkklpqaeemfehlsgklqgfmenksrmlasngpqdihf 2155
Qy 732 WEERRRVRGRADSFIAARMHLVQGDRRFSPWKSGSVVDVIGVFLTQNVYSDHLSASFNSL 791
Db 2156 fqq-----iqelnlemedqgen-----ldtl-----ehlvitelsscgfald 2191
Qy 792 AARPPPKLSSSRREDERNVRSVVVE---DPEGCILNLE-----IPSWQEKVQ---HPS 838
Db 2192 lcqhqrqvnrlkdfetelqktvkerekdasscqelqkrlvrtfkwlketeqsppt 2251
Qy 839 DMEVSGVD-----SGSKQLRDCNSG---IERFNFLEKSIONLEEEVLS--SQDSFDPAI 889
Db 2252 etsvsakeleqiehlkldldwaskgtlveeincnkgtslenlmeitapdsqgktdlte 2311
Qy 890 FQSGRVRGSCSCSDAEFF-----TTRCET--KTVSGTSGSVQF-- 927
Db 2312 lq-----cdmsdvnlkyeklvgvlhergeslqalnlrmeevhkeansvlqleske 2362
Qy 928 -----GSPNLSDEICIQGNER-----PHLYEGSGDVQ--KQETTNVAKKPKDLEXTM 972


```
Db 2363 evlksmdamsptktetvkaqaesnkafaelaeqspkikvkealagllvtypnsgeae 2422
QY 973 NMKDSVCFQPRNDNW-----QTPSSSYEOC-----ATROPHVLDIE-DFG 1014
Db 2423 nwkk1-----qeelnrweratevtvarqqlaesashlacfaaesqlrplwimekelmng 2478
QY 1015 MOGEGIGYSWMSISPRVDRVKNKNVPRFRFGGVSVPREFTQIIIPSTPHELPGMG-LSG 1073
Db 2479 vlgp-----lsldpumlnaqkqv--qfmkfearrqgheqlneaagqiltpgdvs1 2530
QY 1074 SSSAVOEHODDTQHNOQDEMNKASHLOKTFLLDLSNSECLTRQSSSTKONITDGC1PRDR 1133
Db 2531 stsqvqke-----lqslngkwveltdklnsrssqldqalvkst 2568
QY 1134 TAEDVVVDPLSNSSL--QNILVESNNSKEQTAVEYKETNATILREMGKTLADGRKPTSQ 1191
Db 2569 qyqellqlisekvragqrlsvqsaistqpeavkqqlleetseirsdleqlhevkeaqtl 2628
QY 1192 WDSLRKDVEGNEGROERNKNNMDSIDYAIRRASISEISEAIKERGMNMLAVRIKDFLE 1251
Db 2629 cde1svli-geqylkelkrietv-----alplggleadlaadrinriqa 2672
QY 1252 RIVKDHGGIDL-----EWLRESPPDKADYLLSIRGLGKS-----VECVRLTLHLNLF 1301
Db 2673 alastqfqgmfdelrtwldkqsgqakncplsa1erlqsqlqeneefqkslnqhsdgy 2732
QY 1302 PVDTNVGVIAVRMGWVLOPLPESLQLHLELYPVLESIQKFLWPRLCKL----DQRTLY 1357
Db 2733 evlvaegesallls--vppgeekrtlgnqlvelknhweelskktadqrslkdcmqkaqy 2790
QY 1358 ELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPERSTSATIPVP 1417
Db 2791 qwhve-----dlvpwiedck---akmselrvtld----- 2816
QY 1418 PESFPVVAIPMIELPLLEKS-LAGAPSN---RENCEPIIEEPASPGQECTEITESDI 1472
Db 2817 -----pvqlessllrskamlnevekrsllellnsaadllinsseadedgi 2862
QY 1473 ED--AYYNEDPD-----EIPTIKLNIEQFGWTLREHME--RNMELQ-EGDMSKALV--ALH 1521
Db 2863 rdekaginqmdavteeelqakigsleemtqrlrefqesfkniekkvegakhqlEIFdalq 2922
QY 1522 PTTTSIPTPKLNISRLRTEHQVYELPDS-----HRLDGMKREPDP--DPSPYLLAIWT 1574
Db 2923 sqacs-----nknlekitraqevqlqalepqvdylnftqglvedapdgsdasqll----- 2972
QY 1575 PGETANSAPPEQKCGKAGKMKCFDETCSECNLSREANSQTVRGTLIPCRRTAMRGSP 1634
Db 2973 -----hqaevaqge-----flevkqrwns-----gcvmmenkieg1qgf- 3006
QY 1635 LAGTYFQVNFELFADHESSLKPIDVPRDWIWDLPRTT 1670
Db 3007 ----hcrvremf-----sqldlddelgmggaigrdt 3034
```

Search completed: September 27, 2002, 12:54:41
Job time: 10230 sec

THIS PAGE BLANK (USPTO)

; TELEX: 25-3856

```

; INFORMATION FOR SEQ ID NO: 5:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1780 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
US-08-769-309A-5

```

Query Match 2.1%; Score 195; DB 1; Length 1780;
Best Local Similarity 17.8%; Pred. NO. 7.6e-07;
Matches 332; Conservative 262; Mismatches 690; Indels 584; Gaps 84;

Qy	105	ATKPNANWDMSNKSPVTLKSCRKALNFLEDPGDARGDSESEIVQNS-----SGAN	158
Db	23	AEPEPSGGPSAEAAPDPT-----ADPALA-ASDPATKLLQXGOLSTINGVA	69
Qy	159	SFSEIRDAIGTNGSFLDSYSDIKTNGILGAMNOPLE-----VSMGNQPD	203
Db	70	EQDELSLOEGDLNG-----QKALNGOGALNSQBEEVIVTEVQGRDSEVDSESDSK	122
Qy	204	KLSTGAKLARDOQPDLLTRNQOC--QFPVATON-----TQPMENOQAWLMKN--OLIGJP	256
Db	123	EMATKSAAVHIDITDDGOEENRTEIQPSSESNLEELTQ--PTESQANDIGFKVKFVGVFK	181
Qy	257	FGNOQPR-----MTIR-----NQOPLAMG-----NQOPMYLIGTIPRA	290
Db	182	FTVKKDKTEKPDPTVOLLTVKKDEGEAGAAGDHQDFSLGAGEAASKESEPKQ--STEKEP	239
Qy	291	LVSNGOOLGGPOGNKRPIFLNHOTCLPAGNQLYGSPTDMHOLVMSTGGQOHLTKNQOP	350
Db	240	ETLKREQ-----SHAEISP-----PAESQAVVEEC--KEEGEEKOEKEP	276
Qy	351	GSLIRGOOPCVPLIDQOPATPKGFTHLNOMVATSMSSPGLRPHSOVSOPITYLHVESVR	410
Db	277	SK--SAESPSTPVTSEGTSTFKF--FTQGWAGWRKKTSTFRPKDEVEBAS-----	323
Qy	411	ILNGTTTCORSRAPYDSIQDIHOGNKYILSHEISNGCKKALPONSLSIPTPIMAKL	470
Db	324	-----EKKKEQEPEKVDTEED--GKAEVASEKLTASEQAHPQEPESAHEPRLSAEY	373
Qy	471	BEAR-GSKROYHRAMGQTEKHDNLNLAQOIAQSDODVERHNS-----TCVEYLDAAKT	522
Db	374	EKVELPEEGVSGSQSPSEKPAPLATEVF-DEKIEVHOEEVVAEVVHVSTVEERTBEQKT	432
Qy	523	KIOR-----VVOENLHGMPPEVIEE-----DQPTDCARKKNTASTISKA	563
Db	433	EVETAGSVPAEELVGMADAPQAEAPAKELVLUKETCVSGEDTQCADJSPDEKVLUSK--P	491
Qy	564	SKGNSSPVKTAKEKICIVKPTPAK-----KGRAGRKKSVPVPAHAASEIQLWQPTPP	615
Db	492	PEGVSVSEVMSLSOERMKVQGPLKLFTSTGLKLLSGKKQKRGGDDESGEHTQVPA	551
Qy	616	KTPLSRKPKGKGRKSTQDSGKARGPSGELLCC--QDSIAEIIYRMONLYGDKEREQONA	674
Db	552	DSPDSQEOEQGESSASSPEEPE-----EITCLUEKGLAEV-----QQDGEAEQA	595
Qy	675	MVLYKGDG-----ALVPYES-----KKRKP RP-KVDIDDETTRTNWLLMG-----	713
Db	596	-----TSDQEKKRGVTPWASFKKMVTPKKRRVRPSESDKEDLDKYK SATLSSTESTASE	651
Qy	714	-----KQDEKEGDEBKDKK--EKWME-----EER	736
Db	652	MQEEMKGSVEEPPPEPKRKVDTSVSWEALICVSSKKRARRSSDSEGGPKAMGGDHO	711
Qy	737	RVFRGRADSTFARMHLVQDORRFSPMKGSVYDVGIVFLTQNYSDHLLS--SAFMSLAARF	795
Db	712	KADEAGDKETGTDGILLAGSQEHDPGQSSSPSQAG---SPTBEGEGVSTWSEFKRLVT--	766
Qy	796	PPKLLSSREDERNVRSV-----	813
Db	767	PRKKSXKLEKSEKSDSITAGSGVEHSTPDTPEGKEESWVSIKFIPOGRRKKRPDQEOQAP	826

RESULTS 3

RESOL 3
US-08-994-570-5
; Sequence 5, Application US/08994570
; Patent No. 6090929


```

; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-4

```

```

Query Match 2.1%; Score 188.5; DB 2; Length 1435;
Best Local Similarity 18.3%; Pred. No. 1.9e-06;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;

QY 441 ILSHEISNGCKKALPQNSLPTPIAK-----LEEARGSKROYHRAMGQTEKHDLNL 494
DB 299 MLSEHKNNINCKN-IPQELQITQWIKWHGFEFLERDRNSKLPKSKCKNNT----- 350
QY 495 AQIAQSQDVHRHNSSTCVYLD-----AAKTKIQVVOENLHGMPPVEIEIE 543
DB 351 -----LYEACEKECIDPKMYRWIIRSKFEWHTLSKEYETQKVPKNAENY---LKIS 402
QY 544 DDPDYGARK-----GKNATISGASKGNSPVKKTAKEKCIIVKTPAK 588
DB 403 ENKNDKAVSLLNNDCAEYSKYCDCKHTTLVKSVLNGNDNTIK---EKREHIDLDDFSK 459
QY 589 KGRAGRKSVPPAHASEIOLWQPTPKTPLSRKPKGKRKSTQDSKGARGSGELLQ 648
DB 460 FG--CDKNSVD-----TNTKWCKKNPYI-----LSTKD-----VC- 488
QY 649 DSAETIYRQNLVGLDKREQEONAMVLYKGDALVPYESK--KRKPRPKVD-----ID 701
DB 489 -----VPPRQELCLGNIDRIYDKNLLMIKEHILATAIYESRIILKRYKKNDDKEVKII 543
QY 702 DET-----TRIWN-----LLMGKGEKGEDEKDKKKEK-----WVEERR----- 737
DB 544 NKTADIRIDTIGTDYNDLNSRKLKVGKINTNSKYVHRNKKNDKLFDEWKKVKKDVMN 603
QY 738 -----VFRGRADSFARMHLVQGRDRRSPKGSVVDSVIGVFLTQNVSDHLSLSAFAWSLAA 793
DB 604 VISWVFK---DKTVCKREDDIENIPQFFRWFSEWGDY----- 637

```

```

QY 794 RFPKILSSSRREDERNVRSVVVE-----DPEGCILNLNEIPSWOEKVQHPSPDMEVSGVDS 847
DB 638 -----CODKTKMIETLKVECKEPCEDNDCKSKCNSYKWKISKKK----- 677
QY 848 GSKEQLRDCSNSGIERFNFLEKSIQNLLE-----EVLSSQDSFDPALF-----QSCGRVGS 898
DB 678 -----EYNKOAKQOYEQYQGNKMYSEFKSIKPEVLYLKYSEKCSNLF 723
QY 899 CSCSK-----SDAEFPTRC-ETKTV-----SGTSQSVQTSQPNLSDEICLQGNRPHLYE 948
DB 724 EDEFKEELHSDYKNKCTMCPEVDVPLISIRNNEQTSQEAVPPEENTEIA-HRTETFSISE 782
QY 949 G-SGDVQKQ-----ETTINVAQKKPDLKTMNMKDSVCFQKPR 984
DB 783 GPKGNEQKERDDDSLSKISVSPENSRPETDAKDTSLNLLKLGVDVISM---PKAVIGSSP 839
QY 985 NDTNWQTTSPSSSYEQCATR-----QPHVLIDIEDFGM-QGEGLYGSHMSISPRV----- 1031
DB 840 NDNINVTQGDNISGVNSKPLSDVDRPKKELEDQNSDESEETVWNHISKSPSINNGDSS 899
QY 1032 -----DRVKNKNVPRFRFROGGSVPRE 1053
DB 900 GSGSNAVSESSSSNTGLSIDDRNGDTFVRTQDTANTEDVIRKENADKDEKGADEERH 959
QY 1054 FTQIIIPSTPHELPGMGLSGSS---SAVQEHQDDTOHNOQD---EMNKASHLQKTF--- 1103
DB 960 STSESLSSPEEKMLTDNEGNSLNHIEVEKHTSNSDNVQSGGIVNVMNVEKELKDTLENP 1019
QY 1104 ---LDLLNSEECLTRQSTKQNTDCLPRDRTAEDVDPLSNSSLONLIVESNSSNK 1160
DB 1020 SSSLDEGKAHEELSEPNSDDQMSNTPGPLDNTSEETTERISNNEYKVNREDERTLTK 1079
QY 1161 EGTAVEYKETNATILREM-KGTIADGKKPTSQWDSLRLKQVE---GNEGKQ----- 1206
DB 1080 EYEDIVLK---SHMNRSDDGELYDENSLSLTVNDESEDAEAKMKMGNDTSEMHSNHSQHI 1136
QY 1207 --ERNKNMDSID-----YEAIRRASISEISEAIKER 1236
DB 1137 ESDQKNDMKTVGDLGTHVQNEISVPVTGEIDKELKRESKESKIKHAEER 1187

```

```

RESULT 6
US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned

```



```

; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH21.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
;
; US-08-487-826B-4

```

Query Match 2.1%; Score 188.5; DB 2; Length 1435;
Best Local Similarity 18.3%; Pred. No. 1.9e-06;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;

Qy	441	ILSHSEINGNGKALPQNSLPTPIMAK-----LEEARGSKQYHRAWGQTEKHDLNL	494
Db	299	MLSEKHNNINCKN-IQEELQITQWIKEWGEFLLERDSKLPKSKCKNNT-----	350
Qy	495	AQIIAQSGDVERHNSSTCVEYLD-----AAKTKIKVQVENLHGMPPEVIEIE	543
Db	351	-----LYEACEKEGICPCMKYRDWIIRSKFEWHTLSKEYETQKVPKENAENY--	402
Qy	544	DDPTDGARK-----GKNTASISKAGKGNSSPVKKTAKKEKCIIVPKTAK	588
Db	403	ENKDAKVSLLNNCDAAEYSKYCDCKHTTLLVSVLNGNDNTIK-----EKREHIDLDDEFK	459
Qy	589	KGRAGRKSSVPPAHASBIQLWQTPPKTPLSRKPKGKRKSIQDSKGARGSGELLCQ	648
Db	460	FG--CDKNSVD-----TNTKWCKNPVI-----LSTYD-----VC-	488
Qy	649	DSIAEIIYRMQNLYLKGKEREQEQANVLYKGDGALVPYESK--KRPKRPKVD-----ID	701
Db	489	-----VPPRROELCGNIDRIYDKNLMIKEHILAIATYESRILKRYKKNDDKEVCKII	543
Qy	702	DET-----TRINV-----LLMGKGEKGEDEKDKKKEK-----WWEERR	737
Db	544	NKTFADIRIIGGTDYWNLDJNSRKLVGKINTNSKYVHRKNKNDKLFRDEMWNKVKQYWN	603
Qy	738	---VFRGRADSFARHVLVGDRRFPWKGVSVDVSVIGVFLTVQNSDHLSSAFMSLAA	793
Db	604	VISWFK---DKTCKEDDIIENIQFFRFWSEWGDDY-----	637
Qy	794	RFPKLSRSSREDERNRVSVVVE-----DPGCGILNLNLEIPSWQEKVQHPSDMEYSGVDS	847
Db	638	-----CQDKTWIETLKVCEKKEPCEDDCKNSKNSYKEWISKK-----	677
Qy	848	GSKEQLRDCNSGIERNFLEKSTQNLEE-----EVLSSQDSFDPAIF-----QSGR	898
Db	678	-----EEXNQAKQYQEQYQKGNMYSEFKIKPEVLYKKYSEKCSNLNF	723
Qy	899	CSCSK---SDAEPTTRC-ETKTV-----SGTSQSVQTGSPNLSDCLOGNERPHLYE	948
Db	724	EDEFKEELHSDYKNKCTMCPEVKDVPISIIIRNEQTQSQEAYPEENTEITA-HRTETPSIE	782
Qy	949	G-SGDQVKQ-----ETTNAOKKPDLEKTNWKNKDSVCFGQPR	984
Db	783	GPKNEQEKERDDSLSKISVSPNSRPETDAKDTSNLLKLGQDVDISM---PKAVIGSSP	839
Qy	985	NDTNWQTPTPSSSYEQCATR-----QPHVLVDIEDFGM-QGEGLGYSWMSISPRV	1031
Db	840	NDINNVTEQGDNIQSVNSKPLSDVDRPKDKELEQNSDESEETVYNNHLSKSPSINNGD	899
Qy	1032	-----DRVKNKNVPRRFRFGQGSVPRE	1053
Db	900	SGSATVSESSSSNTGLSIDDDRNGDTFVRQDQNTADVIRKKNADKDEDEKGADEPRH	959

Db 63 LMKREMCVSGDDHTQTLTDLSPSEKTLKHPGEGIVSEVMSLSQERIKVQSGPLKLFSS 122
Qy 388 PGLRPHS-----QSQVPTTYLHVESVSRILNLTGTCQSRAPAYDSLQOQIHQ 436
Db 123 SGLKLSGKKQKRGKGDEPEYQHITHS-----PESADEQK 163
Qy 437 GNKYILSHEISNGCKKALPQNSLPTPIAKIL--BEARGSKROYHRAMGOTEKHDLNL 494
Db 164 GESSASSPEEPETCLEKGP---LEAPRMGLKRLKELLRGEKKR-----KDHSLGI 211
Qy 495 AQAQASQDVERHNSSTCVELYDAAKTKTKQVQVQENLHGMPPVIEIDD-----545
Db 212 LOKDGTQETVRRPS-----ESDKEELEKVKKSATLSLSDSTVSEMODEVKTVGBEQK 264
Qy 546 PTGARGKKNFAS---ISKASK-----GNSSPVKKTAKEKC---580
Db 265 PEPKRRVDTVSWEALICVSSKKRKARKASSDIRPRTILGGQOSQSGOQQRSD 324
Qy 581 IVPKTPAKKGRGKKSVPP---PAHASEIQLWQP-----TPPKTPLSRKPKGKGRKSI 632
Db 325 AVPASTQEQDAQGSSSPAGSPSEGEVSTWSEFKRLVTPRKKSCKLEKEAGRTL 384
Qy 633 QDSG-----KARPSGELLICODSIAEIIYRMQNLVLCDK-EREQEQNAM 675
Db 385 VGACPLRSNRVEXNLGFLPLRNSPDG-----GRKGOMGRQSOAT 424
Qy 676 VLYKGDGALVPEYSEKRRKPRKVDIDDETTRIWNLLMGKDEKDEKDKKKEKMWEE 735
Db 425 VEDSG-----DVPVAVPL-----SEYDAVEREKME-----458
Qy 736 RVFRGRADSFIAHMLVQGRDRFSPWKGVSDVIGVFLTONVSDHL--SSSAFMSLAA 793
Db 459 -----AQGNALPSCWCVVSELSKTLVHTVSVAVIDGTRAVTSVEE 501
Qy 794 RPPKLSSS-----REDERNRSVVVED-----PECCIILNLNIPSWQ 831
Db 502 RSPSISASVSTPELHTAGEAMPVEEVTEKDIITAEETPVLTQTLPEG-----549
Qy 832 EKVQHPSDMEYSGVD-----SGSKE-----QLRDCSNS 859
Db 550 -KDAH-DMVITSEVDFTEAVTATSETSEALRTEEVTEASGAETITDMVSAYSQLTDSPT 607
Qy 860 GIER--FNFLEKSTONLEEVLLSODSDPAIFQSCGRVSCSCSKSDAEPTTRCETKT 917
Db 608 TEATPVOQVEGGVLTDEEERQTO-----AILQAV-----ADKVKESQVPATQTVORT 657
Qy 918 VSGTSQVQSGPNLSDBICLQGNRPHLYEGSGDVQKQETTNVAOKKPDLEKTMNWKDS 977
Db 658 GSKALEKE-----EVEEDEVLAKEKEDVMPKGPVQAGAEHLAQ-----699
Qy 978 VCFGQPRNDTNQTTSPS-----SYEQCATRPHVLDIEDFGMGQEGILGYSNMSTSP 1029
Db 700 -----GSETQATPESLEVPENTADVHVATCO--VIKIQOLMEQ-----AVAP 741
Qy 1030 RVDVKNKNVPRFRQGSVPREFTGIIIPSTPHELFCMGLSGSSSAVQEHQDTHQNO 1089
Db 742 ESSETLTD-----ETNGSTP-----LADSDTADGTQQDETIDSO 776
Qy 1090 QDEMKNASHLOKTFDLNLSSECLTROSSTKONITDGLCPDRDAE-----DVPDPLSN 1145
Db 777 DSKATAAQRQSOV-----TEEAATAKEEPSTLPNNVPAQEEHGEPPGRDVLPTQOE 830
Qy 1146 SSLQNILVESNSNKEQTAVEYKETNATILREMKGTLAGKK-----PTSQ--1191
Db 831 LAAAAPVWQKTEVQGEQEDV-----LDGEKVKEEQEVFVHSGPNSQKA 875
Qy 1192 -----WDSLKRDVEGNEGQERKNMWSIDYEA-----IRRASISESEAIKERGMNM 1241
Db 876 ADVTYDS---EYMGVAGQCEKSTEVQSLSEEGEMETDVEKRETRPEQVSEEGEQT 932
Qy 1242 LAVRIKDFLERIVKHGGIDLEWLRESPPDKAKDYLLSTIRGLKSVCEVRLTLTHNL--1299
Db 933 AA-----PEHERNKGKPVLTILDMPSSE--RCKALGSLGG 964

Qy 1300 --APPVDTNVGRIAVRMGWVPLQPLPESLQHLLELYPVLESQKFLWPRCLKLDORTLY 1357
Db 965 SPSLPDQDKAGCIEVQV-----QSLDTTWTQ--TAEAVEKVI-----999
Qy 1358 ELHYQLITFGKVFCTKSRPNCNACPMRGECRHFAVASARLALPAPEERSL TSA-----1412
Db 1000 -----ETVVISSE---TGSPSEC-----VGAHL-LPA-EKSSATGGHWTQ 1034
Qy 1413 ---TIPVPESFPVPAIMTELPLPLEKSL-----ASGAPSNRENCEPIIEEPASP---1460
Db 1035 HAEDTVPLGPES-QAESIPILVTPAP-ESTLHPDLOGEISASQORSEERSEEDKPDAGPDAD 1092
Qy 1461 GOECTEI-----TESDIEDAYNEDDPDEIPTIKLINIQFGMT-----1497
Db 1093 GKSTAIKVLKAEPEILELESKNKIVLNVIQTAVDQFARTETAPETHAYDSQTVQVAM 1152
Qy 1498 -----LREHMERNMELOEGDMSKALVALHPTTTISI-----1527
Db 1153 RLDSREPNRCWTKMKVAKMKHPVQPPREDIQVLTVLEAWLSSEMLAALAVESAGVKYSIE 1212
Qy 1528 ---PTPKLKNISRLRTEH-----QVYELPDSHRLLDGMDKREPDDPSPYLLAIWTPGETA 1579
Db 1213 KLPPQPKDQK-----EHAADGPOLQSLAQAEVYSGNLTRESPTNGPKL-----1256
Qy 1580 NSAPPPKQCGGK--ASGKCMCFDETCSE-----CNSLRE-----ANSQTVRGT 1620
Db 1257 ---TEERCPCQLRSRKKCLPSQSKRTRPRQKRTCSRQRETWQNPKNMLVAHCTSRVP-1310
Qy 1621 LLIPCRTAGMSFPPLNGTYFQVNFELFADHE 1650
Db 1311 ---ECENKSONKMLLLGPWTKISEPMKRSRE 1337
RESULT 8
5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VESICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VESSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO:2
; LENGTH: 2409
5180808-2
Query Match 2.08; Score 185.5; DB 6; Length 2409;
Best Local Similarity 17.8%; Pred. No. 8.2e-06;
Matches 315; Conservative 234; Mismatches 624; Indels 597; Gaps 81;
Qy 103 ESATKKPKNVGDMSNKSPEVTLKSCRKALNFDLENP--GDARQGDSESEIVQNSGANSF 160
Db 403 EEEEECANADVTT-TPSQVQYINGKHLVTVPKDPEAEARRGQFESVAPSONFSDSSE 461
Qy 161 SEIRDAIGTNGSFLDSVSQIDKTNGLGMNQ-----LEVS-----M 198
Db 462 SDTHPFV-----IAKTE-LSTAVQPNNESTETTESLEVTWKPTETPSETSEHS 507
Qy 199 GNQDPKLT-----GAKLARDQOPDLLTRNQOQFPVATQNTQFPHENQ-Q 243
Db 508 GGEQVFTVPFHEEFESGTAKKGAESVTERDTEYVGHQAHEHTEPVSL-----FPBESSGE 563
Qy 244 AWLQMKNLQIGF-----PFGNQOPRMTIRNQOCPCLA-----MGNOQPMYLCITPRPA 290
Db 564 IADQESOKIAFARATEVTFFGEEVEKSTSVITYTPIVSSASAYSEEAATLIGNPWP 623
Qy 291 LVSGNQQLGGQGNKRPFIHLNHTQCLPA-----GNOLYGSPTDMHOLVMTSGQQ 340
Db 624 DLLSTKESWVEATPRQVVVELSGSSSIPITEGSGEAEDEDTMTMTVTDLSQ-----674

REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9672
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)-546-4737
TELEFAX: (619)-546-9392
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2441 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-194-468-2

Query Match 2.0%; Score 183; DB 1; Length 2441;
Best Local Similarity 17.5%; Pred. No. 1.4e-05;
Matches 342; Conservative 222; Mismatches 665; Indels 708; Gaps 84;

Qy 5 MDSSAVNATEAQDNGSRQVDFDLNKT-----PQOKPSKRKRKPKMPKVVECKPKRK 59
Db 287 MGATGVNPLASKOSWNSLPAFPDPTDIKNTSVTVPNMSQLQTSVGIPTQAIATGPTAD 346
Qy 60 P--RKPAPLKVVEGPKPKRAATQEKVKSETGSAKKKNLKESAT----KKPANVG 113
Db 347 PEKRLIQOQLVLLHAKHCCORQOANGEVACSLPHCRTWKVNLNHNHMHCOAPRACQVA 406
Qy 114 DMSNKSPEVT-LKSKRC---ALNFDLENPGDARQD----SESEIVQNSGANGSFSEIRD 165
Db 407 HCASSRQIISHWKNTCHDCVCLPLKNASDKRQOITILGSPASGIQNTIGS-----V 459
Qy 166 AIGTNGSFLDSVSQIDKTN-----GLGAMNQPLEV----- 196
Db 460 GAGQONATSLSNPNIDPSSMORAYAAALGPLYMNQPTQLQPVQGPQAPPAHQOQRT 519
Qy 197 --SMGNQPKLSTGAKLARDQDPLLTNRQOQCPVATQNTQFMENQOANLQKNQILG 254
Db 520 LNALGNPMSPVAGG-IITDQPPMLISEALPTSLGATN---PLMND-----G 564
Qy 255 PFGNQOQPMRTIRNOQPCLAMGNOQPMY-----LIGTPRPAL----- 291
Db 565 SNSGNIGSLSTIPTAAPSTGVRKGWEHVTQDLRSHLVHLKVAIPTPDPAALKDRR 624
Qy 292 -----VSGNQQLG 299
Db 625 MENLVAYAKKVEGDMYESANSRDEYHLLAEKIYKIQELEKEERKRLHKGILGNQAL 684
Qy 300 GPQGNKRIPIELNHQCLPAGNOLYGSPTDMHOLYMTSGQQHGLLIKNQOQSLIRGOOP 359
Db 685 PASGAQPPVIPPAGSVRPPNGPL---PLPVNRMQVSQG-----MNSFNPMSLGNVQLP 734
Qy 360 CVPLDQOQATPKGFTLHNQMV-ATSMSS-PG--LRPHSQSQVQPTTYLHVESVSRLNGT 415
Db 735 -----QAPMGPRASPWNHVSQNMASVFGMATSPSRMPQPPN-----MMGT 777
Qy 416 TGTGORSRAPAYDSLQDIHOGNKYILSHEITSNGCKKALPQSSSLPTPIAKLEFARG 475
Db 778 HANNITMAQAPTQNOF-----LPQN----- 796
Qy 476 SKROYHRAVGOTERHDLNLAQIAQSODVERHNSSTCVE-----YLDAAKTKIKQVVQE 530
Db 797 ---QFPSSGAMSVNSVGMGPAAQAGVSGOQEPGAALPNLMLAPQASQLPCPPVPTQS 853
Qy 531 NLGHMPPEVIEIDDP-----TDGARKGNKTASISKASK 565
Db 854 PLHPTPPASTAAGMPSLQHTAPQMTPPQAPATQPTPVSSSGTPTPTPGSVPSAQT 913
Qy 566 GNSSPVKKTAKKCIYPKTPAKGRAGRKKSVPPPAHASEIQWLQPT-----PPKTPLS 620
Db 914 QSTPTVQAAQAQVTPQPTQTVQ-----PPSVATPQSSQOQPTPVHTQPPGTPLS 963
Qy 621 RSK-----PKGGRKSIQDSGRARGSGELLQCDSTAEIIRWQNLVLDGKEREQEQN 673
Db 964 QAASIDNRVPTPTVSTAETSSQOQPGDVPML-----EMKTEVQT---DDAEPE----- 1010

Qy 674 AMVLYKGDGALVPYVESKRRKPRPKVDIDDETRINLLMGKDEKDEKDKKKKKWWE 733
Db 1011 -----PTES-KGEPSEMEED-----LOGSSQVKEETDTTEQKSEPMEV 1049
Qy 734 EERRVFRGRADSFARMHLVQDGRFRFPWKGSVVDSVIGVFLTONVSHLSSAFMSLAA 793
Db 1050 EEKK-----PEYKVEAKEEEEESSNDTASQSTSP 1078
Qy 794 RFPPLKSSSREDER-----NVRSVVVEDPEGCIL-----NLNIPSWQEKVQHPDSMEV 842
Db 1079 SQPRKKIKFPEELRQALMPTLEALYRODPESLPFRQPVDPOLLGIPDYFDIVKKNPMDLST 1138
Qy 843 --SGVDSGSKEQ-----LRDCNSGIERFNFLKSIQNLEEEVSSQDS 884
Db 1139 IKRKLDTGOYEPWQYVDDVRLMFNNAWLYNRKTSRV--YKFSKLAEPFQEI----- 1190
Qy 885 FDPALFOSGCRVSGSCSKSDAEFTTTCETKTVTSGTSQSQTGSPNLSDBIC-----L 938
Db 1191 -DP-VMQS---LGYCCGRKRYEFPQTLCCYQKOLCTIPRDAAYYSQNRVYHFCGKCFTEI 1245
Qy 939 QGNERPHLYEGSGDVQKQETTNAQKPDLEKTMN----- 973
Db 1246 QG-ENVTL---GDDPSQPTTIS--KQOFEKKNDILDPPEFVDCKECGRKMHOICVLH 1298
Qy 974 ---WKDS-VC-----FGQPRNDTNW-----OTTPSSSY---EQCATRQPHVLDIED 1012
Db 1299 YDIIWPSCFVDCNCLKKTGRPRKENKFSAKRLQTLRGNHLEDRVNFKLRQNHPEAGEV 1358
Qy 1013 F-----GMOGELGYSWMSTISPRVDVKNKNVPRR-----FFRQGGSVPR 1052
Db 1359 FVRVVASSDKTVEKPGMKSRFVDSGEMS-----ESFYRTKALFAFEEDIGVDV 1408
Qy 1053 EFTGOIIPST---PHELPGMG-----LSGSSSAVOE--- 1080
Db 1409 CFFGMHVDOTALIAPHQIOGCVYISYLDISIHFPRCLRTAVYHEILLGYLEYVKLVYV 1468
Qy 1081 ---HODD---TOHNQOQDEMKNASHLQ-----KTFDLNLSSECLTROSS 1119
Db 1469 TAHIWACPPSGDDYIFHCHPDQIKPKPKRLQEWYKMLDKAFAERIINDYKDFIKQAN 1528
Qy 1120 TKONITDGCCLPRDRTAEDVDPLSNSSLQNLILVESNSNKEQTAVEYKFNATILRMK 1179
Db 1529 -----EDRLTSAKELPYEFDFWPNVLEESIKELQEEERKKEESTASPTPE 1577
Qy 1180 GTLADGKKPTSQWDSLRKDVESNEGROERNKNMDSIDYEAIRRASISEIETAKER--- 1236
Db 1578 GSQDSK-----NAKKNNKATNKK-----SSISRANKKPSM 1611
Qy 1237 -GMNMLAVRIKDFLER-----IVKDHGGIDLEWLRESPPDKAKDYLLSTIRGLKLSVE 1289
Db 1612 PNVSNLQSLQYATMEKHKEVFFVHILHAG--PVISTQPPIVDPDPLLS----- 1658
Qy 1290 CVRLLTLHNLAFPDVTNVGR-----IAVRMGWVPLQPLPESLQLHLLLEYLVLESIOKFL 1344
Db 1659 -----CDLMDGRDAFTLTARDKHW-----EFSLSRRSK 1686
Qy 1345 WPLCKLDQRTLYELHQLITFGK---VFCUKSPNCPNACPMRGCERHFASAYASARLAL 1401
Db 1687 WSTLC---MLVELHTQ---GQDRFVY-----TCNECKHHVETRWCHTVCEDYDLCI 1731
Qy 1402 PAPERSLTSATIPVPPPEPPVAIPMIELPLPL-EKSLASGAPSNRENCEPIITEEPASP 1460
Db 1732 NCYNTKSHTH-----KMKWGLGLDDESSQSGEPQSK-----SP 1765
Qy 1461 QOECTETI-----ESDIEDAYINEDDPEIPTIKLINIQFGMTLREHMERNMELQEG----- 1511
Db 1766 -QESRRLSIQRCISQSLVHACOCRNANGSLPSC-----QMKRVVQHTKQCKRK 1812
Qy 1512 -----DMSKALVAL-----HPTTITSIPTP---KLK-NISRLRTEHOVVELPDSHRLLDG 1556
Db 1813 TNGGCPVCVKQLIALCCYHAKHCQENKCPVCPCLNIKHNVRQOQIOHCLQQAOLMRRRMT 1872


```
Db 1053 EVG-----SSINEIGSSDENIAELG-RNRGPKLNAMRL---GVLQPEVYKQSLPSGNC 1103
Qy 692 KRPRKVIDIDETRIWNL-----LMGKDEREGD 721
Db 1104 KHPEIKQVEYEVQTVNTDESPYLI SDNLEQPMGSHASQVSETPDDLLDDGEIKEDT 1163
Qy 722 E--BKDKKKKWEERVF-----RG---RADSFIA RMHLVQDGRFSPWKGSVVDVSI 771
Db 1164 SFAENDIK-----ESSAVFSKVQSGELSRSPFTHTHLAQYRGA-----1206
Qy 772 GVFLTNVSDHLSSAFMSLAARPPKLSRSDREVRNRSVVVEDPE-GCILNL-----N 825
Db 1207 -----KKLESS---FENLSS---BDEELPCFQHLLFGKVN 1235
Qy 826 EIPSQWQKVQHPSPMEYSGVDGSKG-----OLRDCNSNGIERNFLEKSIQNLLEEV 878
Db 1236 NIPS--OSTRH-STVATECSKNTENLLSKNSLDCSNQVI-----LAKASQ---EHH 1284
Qy 879 LSSQDSFDPALFQSGRVCSCSKSDAEFPFTRCTKTVSGTSQSVQGTGSPN-----LSD 934
Db 1285 LSEETKCSASFSS-----QCSELEDLTANTNTQDPFLIGSSKQMRHQSOGVGLSD 1337
Qy 935 EICQGNRPHLYGSGDVQKQETTNVAQKPDLEKTMNWKDSVCFQOPRNDTNWQTTPS 994
Db 1338 KELVSDDEE---RGTG---LEENN--QEQSMDSNLGEAAGC-----BSET 1376
Qy 995 SSYEQCATRQPHVLDIEDFGMGEGSLGYSNMSISPRVDRVKNKNVPRFFRQGSVPREF 1054
Db 1377 SVSEDCS-----1383
Qy 1055 TGQIIPSTPHLPGMGLSGSSSAVQEHQDDT-QHNOODEMKNKASHLOKTFDILLNSSEEC 1113
Db 1384 -----GLSSQSDILTTQORDTMQHN-----LIKQOEMAELEAVLEQH 1421
Qy 1114 LTROSSTKONTDCLPRDRTAEDVDPLSNSSQLNVLVSNSSNEKEQTAVEYKFNAT 1173
Db 1422 GSQPSNYSPT-----ISDSSALEDLRNPEQSTSEKAVILTSOKSEYP 1464
Qy 1174 ILREMGKTLADGKPKTSQWDSLRKDVEGNEGRNKNMDSIDYEAIRASISEAISEAI 1233
Db 1465 ISQNEGLSAD-KREVSADSTSKNKE--FGVERSSPSKCPSLD-----DRWYMHSCSGSL 1517
Qy 1234 KERGM-NMMLAVRIKDFLERIVKDHGGIDL 1262
Db 1518 QNRNYPQSEELIKVVDVEEQLESGPHDL 1547

RESULT 13
US-08-798-691-6
; Sequence 6, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvarez, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; US-08-798-691-6
```

```
Query Match 2.0%; Score 182; DB 1; Length 1863;
Best Local Similarity 18.4%; Pred. No. le-05;
Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

Qy 20 DGSRODVLER---DLNKTPOQPKSKRKRKEMPKVYVVGKPKR-----58
Db 219 DSAKKAACEFSETDVTTEHQPSNNDLNTTEKRAARHPEKYGSGSVSNLHVEPCGTNT 278
Qy 59 -----KRPKPAELPKVYVVGKPKR-----KRPKPAATQEKVKSKETGSAKK 98
Db 279 HASSLQHENSSLLLTDRMNMVKEAFCNKSQKGLARSHQHNWAGSKETCNDRRTPSTEK 338
Qy 99 K-----NLKESATKKPANVGDMNSKSEVTLKSCRKALN-----FDLEN 137
Db 339 KVDLNADPLCERKEWNKQKLPCESENPRDTEDEV---EWITLNSSIQKVNEMFSRDELLG 394
Qy 138 PGDARQDSESEI-----VQNSSGANS-----FS 161
Db 395 SDDSHDGESENAKADVLDVLNEVDYSGSSEKIDLLASDPHEALICKSERVHSKSVES 454
Qy 162 EIRDAIGGTN-----GSFLDSVSQIDK---TNGLGAMNQPLEVSMG 199
Db 455 NIEDKIFGKYRKASUPNLSHVTENLIIGAFVTEPQIIQERPLTNLKRKRKP---TSG 511
Qy 200 NOPDKLSTGAKLARDQOPDLLTR--NOQCQFPVATQNTQPMENQOAWLQMKNLIGFPF 257
Db 512 LHPEDFIKKADLAVOKTPEMINOCTNOTEQNGQVMNTTNSGHENKTKGDSIQNE-----565
Qy 258 GNOQPRMTIRNOOPCLAMGNOQPMYLLIGTPRALVSGNOOLGGPGQGNKRPIFLN-HOTCL 316
Db 566 -----KNPNPIESLEKESAFKTAEPISISSIS-NMELE-----LNHNSKA 605
Qy 317 PAGNQLYGSPTDMH---OLVMS-----TGGOQHGLLIKNQOPGSLIRG 356
Db 606 PKNRLRRKSSSTRHIALELVVSRNLSPNPCTELQIDSCSSSEIEIKKKYNQMP---VRH 662
Qy 357 QQPCVPLIDQOPAT-----PKGFT-----HLNQMV 381
Db 663 SRNLQMEGKEPATGAKKSNKPNBQTSKRHSDTTFPELKLTNAPGFSFTKCSNTSELKEFV 722
Qy 382 ATSM-----SSPGLRPHSQSQVP-TTYLHVES 407
Db 723 NPSPREEKEKLETVKVSNNAEDPKDMLMSGERVLOTERSVSSSISLVPGTGYGTQES 782
Qy 408 VSRILNCTTGTQCR-----SRAPAYDSLOODIH-----QGNKYILSHSISNGNG 451
```


Db 783 ISLEVTGLKAKTEPNKCVSQCAAFENPKGLIHGCKDNRRNDTEGPKYPLGHEV----- 837
Qy 452 CKKALPONSLPTPIMAKLEA-----RGSRRQ-----YHRA 483
Db 838 -----NHRSRTSI--EMESELDQYLONTFKVSRQSFALFSPNCAAECAATSAH 888
Qy 484 MGOTEKHLNLAQIAQODVERHNSSTCYEYLDAAKTKIKYVQBNL-HGMPPEYIEI 542
Db 889 SGLKQSPKVTPECEQKEENQGNES-----NIKPVQTVNITAGFP--VVGQ 934
Qy 543 EDDPTDGR---KCKNTASIS---KGASKGNSSPVKTAKEKCIVP-----KTPA 587
Db 935 KDKVVDNAKSIKGGSRFCSQFRGNETGLITPNKGLLQNPYRIPPLPKSFVTKC 994
Qy 588 KKGKGRKKSPPPAHASEIQLOMPTPKT--PLSRSKP-----KKGKRSIODSGKARG 640
Db 995 KKNLL--EENFEHSMSPEREMGNENTPSTVSTISRNIENRFKGAASSNINEVGSTN 1052
Qy 641 PSGELLQDSIAEIIYRMQNL--LGDKEREQONAMVLYKGDAIYVYESKK----- 691
Db 1053 EVG-----SSNETGSSDENIQAEIG--RNRGPKLNAMLRL---GVLPQEVYKQSLPGSNC 1103
Qy 692 RKRPKYDIDDETRINWL-----LMKGDEKEGD 721
Db 1104 KHPEIKKQEEYEVQTVNTDFSPYLISDNLEQPMGSSHASQVCSETPDDLDDGEIKEDT 1163
Qy 722 E--EKDKKKKEKWEERRVF-----RG---RADSFIAHMLVQGDGRFSPWKGVSVDVI 771
Db 1164 SFAENDIK-----ESSAVSKSVQRELSRSPFTHLQAYRRGA----- 1206
Qy 772 GVFLTQVNSDHLSSAFMSLAAREPPKLLSSREDERNVRSVVVDPE--GCILNL-----N 825
Db 1207 -----KKLESS---EENLSS---EDELPCFQHLLEKVN 1235
Qy 826 EIPSWQKVQHPDMEVSGVDSGSKE-----QLRCSNSGIERFNFLEKSIQNLSEEV 878
Db 1236 NIPS--QSTRH--STVATECLSKNTEENLLSLKNSLDCSNQVI-----LAKASO---EHH 1284
Qy 879 LSSQSDPDPAIFQSGRVSQSGSKSDAEPPTTCETKTVSGTSQSVQGTSPN-----LSD 934
Db 1285 LSEETKCSASLFSS-----QCSELEDLTANTQDPFLIGSSKQMRHOSQGVGLSD 1337
Qy 935 EICLOGNERPHLYGSGGVOKQETTNVAQKPDLEKTMNMKDSVCFQOPRNDTNWQTTPS 994
Db 1338 KELVSDDEE---RGTG---LEENN--QEQSMDSNLGEAASGC-----ESET 1376
Qy 995 SSVQCATRQPHVLDIEDFGMQGELGYSNMSISPRVDRVKNKNVPRFFRQGSVPREF 1054
Db 1377 SVSEDCS----- 1383
Qy 1055 TGOIIPSTPHELPGMGLSGSSSAVQEHQDDT-QHNOODEMKNKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILTTQORDTMQHN-----LIKLOEAMAEVLEAVLEQH 1421
Qy 1114 LTROSSTQNITDGLCLPRDRTAEDVDPLNSNLSQNLIVESNSNKEQTAVEYKETNAT 1173
Db 1422 GSQSNVYPSI-----ISDSSALEDLRNPQESTSEKAVILTSQKSSEYP 1464
Qy 1174 ILREMKGTLDGKKPTSQWDSLRKDXVEGNEGRQERNKNWDSIDYEAIRASISEISEAI 1233
Db 1465 ISQNPGLSAD-KFEVADSSTSKNKE--PQVERSSPSKCPSLD---DRWMHSCSGSL 1517
Qy 1234 KERCM--NMMLAVRIKDFLERIVKDHGGIDL 1262
Db 1518 QNRNYPQEEELIKVVDVEEQOLESQPHDL 1547

RESULT 14
US-08-825-487A-2
; Sequence 2, Application US/08825487A
; Patent No. 6048689
; GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.
APPLICANT: White, Marga B.
TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue., N.W.
CITY: Washington,
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,487A
FILING DATE: 28-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US98/060002
FILING DATE: 26-Mar-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 05371.0012.999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8100
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-2

Query Match 2.0%; Score 182; DB 3; Length 1863;
Best Local Similarity 18.4%; Pred. No. 1e-05;
Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;
Qy 20 DGSRODVLEF---DLNKTPOQKPSKRRKEMPKVVVEGPKR----- 58
Db 219 DSAKKAACESETDVTTEHQPSNNDLNTEKRAAEHPEKYOGSSVSNLHVPCGTNT 278
Qy 59 -----KRPKPAELPKVVVEGK-----KRPKPAATQEKVKSKETGSAKK 98
Db 279 HASSLOHENSLLLTDKRMNVKAEFCNKSQPCQLARSQHNRWAGSKETCNDRTPTSTEK 338
Qy 99 K-----NIKESATKKPANVGDMKNSKEVTLKSCRKALN-----FDLEN 137
Db 339 KVDLNADPLCERKEWNKQKLPCSENPRDTEV-----PWITLNSSIQKVNWFNFSRDELLG 394
Qy 138 PGDARQGDSESEI-----VQNSSGANS-----FS 161
Db 395 SDDSHGESESNAKVADVLDVLNEVDYSGSEKIDLLASDPHEALICKSERVSKSVES 454
Qy 162 EIRDAIGTGN-----GSFLDSVSGIDK---TNGLGAMNQPLEVSMG 199
Db 455 NIEDKIEGKYRKASLPNLSHVNTENLIIGAFVTEPQIIQERPLTNLKKRRRP---TSG 511
Qy 200 NQPKLSTGAKLARDQOPDLLTR--NOOCOPFVATQNTQPFMENOQAWLQMKNLGIFPF 257

Db 512 LHPEDFIKKADLAVQKTPPEMINOGTNOTEQNGQVMNITNSGHENKTKGDSIQNE----- 565
QY 258 GNOQPRWTIRNOOCLAMGNOQPMYLGTRPALVSGNOQLGGPQGNKRPIFLN-HOTCL 316
Db 566 -----KPNPIESLEKESAFKTAEPISISSIS-NMELE-----LNIHNSKA 605
QY 317 PAGNQLXGSPDMMH-----OLVMS-----TGGOQHGLLIRNOQPGSLIRG 356
Db 606 PKKNRLARKSSTRIHHALELVVSRNLPNPPNCTELQIDSCSSSEIEKKKKYNQMP---VRH 662
QY 357 QOCVPLIDQOPAT-----PKGFT-----HLNQMV 381
Db 663 SRNLQMEGKEPATGAKSKNPNQTSKRHSDTFPELKLTNAPGSPFTKGSNTSELKEFV 722
QY 382 ATSM-----SSPCLRPHSOSQVP-TTYLHVES 407
Db 723 NPSLPREKEKEKLETVKVSNNADPKDMLSGERVLOTERSVSSSISLVPDGTDTGYQES 782
QY 408 VSRILNGTTGTCQR-----SRAPAYDSLOQDIH-----QGNKYILSHEISNGNG 451
Db 783 ISLLEVSTLCKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDETEGFKYPLGHEV----- 837
QY 452 CKKALPONSLPPIMAKLEA-----RGSKRQ-----YHRA 483
Db 838 -----NHSRETSI--EMBESELDQYLOLQTFKVSRRQFALFNSPGNAEECATFSAH 888
QY 484 MGOTEKHDLNLAQIAQSDQVERHNSSTCVIELDAAKTKIKVQVQENL-HGMPPEVIEI 542
Db 889 SGSLKQSPKVTPECEKEENQGNES-----NIKPVQTVNITAGFP---VVGQ 934
QY 543 EDDPTDGAR---KGKNTASIS---KGASKGNSPVPKTAKEKCIVP-----KTPA 587
Db 935 KDKPVDNAKCSIKGSRFCLSQSPRGNETGLITPNKHGCLLONPVPPLFPKSFVKTKC 994
QY 588 KKGKGRKKSVPVPAHASETLQWPTPKT--PLSRSKP-----KGKGRKSQDSCARG 640
Db 995 KKNLL--EENFEHSMSPEREMGNENIPSTVSTISRNIRENFVKGASSNINEVGSSTN 1052
QY 641 PSGELLCQDSTAETIYRMQNLV--LGSKEREQONAMVLYKGDGALVPYESKK----- 691
Db 1053 EVG-----SSINEIGSSDENIQALG--RNRGPKLNMLRL--GVLOPEYVKQSLPQSN 1103
QY 692 KRPKPVVIDDETTRIWL-----LMGKGRDEKED 721
Db 1104 KHPEIKKQYEEVVQTVNTDFSPYLSIDNLEQPMGSSHASQVCSQVSETPDDLDDGETKEDT 1163
QY 722 E--EKDKKKKEKWEERRVE-----RG--RADSFIAHMLVQDREFSWKGVSVDVSI 771
Db 1164 SFAENDIK-----ESSAVFSKSVQRGELSRSPFTHTHLAOGYRGA----- 1206
QY 772 GVFLTQNVSDHLSSAPMSLAARFPKPLSSRREDERNVRSVVVEDPPE-GCILNL-----N 825
Db 1207 -----KLESS---EENLSS---EDELPCFQHLFLGKVN 1235
QY 826 EIPSWQKVOHPSDMEVSGVDSGKE-----OLRDCNSNGIERFNFLEKSIQNLDEEV 878
Db 1236 NIPS--QSTRH-STVATECLSKNTEENLLSLKNSLNCNQVI-----LAKASQ---EHH 1284
QY 879 LSSQDSFDPALFOSGCRVSGSCSKSDAEPPTTCETKTVSGTSSQVQGTSPN---LSD 934
Db 1285 LSETKCSASLFS-----QCSLEDLFTANTQDPFLIGSSKOMRHOSESQGVCLSD 1337
QY 935 EICLQGNRPHLYEGSDVQKQETTNVAQKPKDLEKTMNKNKDSVCFQGPQRNDTNWQTTPS 994
Db 1338 KEIVSDDDEE---RGTG---LEENN--QEEQSMDSNLGEAASGC-----ESET 1376
QY 995 SSVBQCATRPHVLIDEDFGMQGEGLOYISWMSISPRVDRVKNKNVPRFRFGQSVVPREF 1054
Db 1377 SVSDECS----- 1383
QY 1055 TGOIIPSTHELPGMGLGSSSAVQEHODDT-OHQOODEMKNKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILTTQOORTMOHN-----LIKQOEMAELEAVLEQH 1421

QY 1114 LTRQSSSTKQNTDGLCLPRDRTAEDVVDPLSNSSSLQNLVSNSSNKEQTAVYKETNAT 1173
Db 1422 GSOPSNYSYPSI-----ISDSSALEDLRNPEQSTSEKAVLTQSKSSEYP 1464
QY 1174 ILREMGKTLADGKKPTQSONDSLKRDVNEGROERNKNNMDSIDYAIRRASISEISEAI 1233
Db 1465 ISQNPGLSAD-KFEVSADSSTSKNKE--PCVERSPPSKCPSLD----DRWYMHSCGSL 1517
QY 1234 KERGM-NMMLAVRIKDFLERIVKDHGIDL 1262
Db 1518 QNRNYSQSEELIKVVDVEEQOLEESGPHDL 1547
RESULT 15
US-08-825-487A-6
; Sequence 6, Application US/08825487A
; Patent No. 6048689
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: White, Marga B.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue., N.W.
; CITY: Washington,
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,487A
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/060002
; FILING DATE: 26-Mar-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Albert P. Halluin
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOC/KET NUMBER: 05371.0012.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8100
; TELEFAX: 650-463-8400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-08-825-487A-6

Query Match 2.0%; Score 182; DB 3; Length 1863;
Best Local Similarity 18.4%; Pred. No. le-05;
Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

QY 20 DGSRODVLPE---DLNKTPOQKPSKRKRKMPKVWVGKPKR----- 58
Db 219 DSAKKAACESEYDVTNTEHHQFSNNDLNTERAAAEHPEKYQGSSVSNLHVPCGTNT 278

QY 59 -----KPRKPAELPVVVEGRP-----RRKPKAAATQEKVASKETGSACK 98
DB 279 HASSLOHENSLLTLTKDRMVEKAECNKSQOPGLARSHQHNRAWSKETCNDRRTPSTEK 338
QY 99 K-----NLKESATKPPANVGDMSNKSPEVTLKSCRKALN-----FDLEN 137
DB 339 KVDLADPLCEKKEWKKQLPCSENPRTEDV-----PWITLSSIQKYNEWFSRDELLG 394
QY 138 PGDARQGSSESI-----VONSSGANS-----FS 161
DB 395 SDDSHGSESNKAVDVLVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSYES 454
QY 162 ETRDAIGTN-----GSFLDSVSOIDK-----TNGLGAMNQPLEVSMG 199
DB 455 NIEDKIFGKTYRKASLPNLHVTEENLIIGAFVTEPOIIOERPLTNKLRKRPP-----TSG 511
QY 200 NOPDKLSTGAKLARDQOQDPLLTR--NOOCQFPVATQNTQFPMENOQALQMKNLQIGPFF 257
DB 512 LHPEDFIKKADJAVOKTPEMINQGTQEQNGQVNMNITNSGHENKTKGDSIQNE-----565
QY 258 GNOQPRMTRINQOPLAMGNOQPMYLIQTPRPALVSGNOOLGGPOGQNKRPFLN--HOTCL 316
DB 566 -----KNPNPIESLEKESAFKTAEPITSSIS--NWELE-----LNHNKA 605
QY 317 PAGNOLYGSPTDMH-----OLVMS-----TGGQOHLIKKNOQPGSLIRG 356
DB 606 PKKNRLRRKKSSTRHIALELVVSRNLSPNCTELQIDSCSSSEIEKKKKYNQMP--VRH 662
QY 357 QOPCVPLIDQOPAT-----PKGFT-----HLNOMV 381
DB 663 SRNOLMEGKEPATGAKKNKPNNEQTSKRHSDTPELKLNTNAPGSFTKCSNTSELKEFV 722
QY 382 ATSM-----SSPGLRPHSQSOVP--TTYLVHVES 407
DB 723 NPSLPREKEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGTDIGTOES 782
QY 408 VSRILNGTGTQCR-----GRAPAYDSLQODIH-----QGNKYILSHEISNG 451
DB 783 ISLLEVSTLGRAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETFKYPLGHEV-----837
QY 452 CKKALPONSLUPTPIMAKLEA-----RGSKRQ-----YHRA 483
DB 838 -----NHSRETSI--EMEESELDAQYLQNTFKYKRSQSFALFSPNGNAEBEATFSAH 888
QY 484 MGOTEKHDLNLAQQAQODYVERHNSSCTVEYLDAAKTKIKQVVOENL--HGMPEVIEI 542
DB 889 SGLKKQSPKVTFECEQKEENQKNES-----NIKPVQTVNITAGFP--VYQ 934
QY 543 EDDPTDGAR--KGKNTASIS--KGASKGNSPPVKKTAEKEKCIVP-----KTPA 587
DB 935 KDKPVDNAKCSIKGSRFCLSQFRNETGLITPNKHGLLQNPYRIPPLFIKSPVKTKC 994
QY 588 KKGRAGRKSPVPPAHASEIQWLQWTPPKT--PLSRSP-----KGGRKSQDSGKARG 640
DB 995 KKNLL--EENFEHMSPERMGENIPSTVSTISRNIRENVFKGASSNINEVGSSTN 1052
QY 641 PSGELLCODSTAEIIRYMNLY--LGDREREQENAMVLYKGDGALVPYESKK-----691
DB 1053 EVG-----SSINEIGSSDENIQAEIG--RNRGPKLNAMELRL--GVLQPEVYKQSLPGSNC 1103
QY 692 RKPRPKVDIDDETTRIWNL-----LMGKGDEKRGD 721
DB 1104 KHPEIKKQYEBEVQVTVNTDFSPYLLISDNLEQPMGSSHASQVCSETPDDLLDDGEIKEDT 1163
QY 722 E--EKDKKKEKWEERRVF-----RG--RADSFIARMHLVQGRFRSPWKGVSVDVYI 771
DB 1164 SFAENDIK-----ESSAVFSKVQGELSRSPPFTTHLAQQYRRGA-----1206
QY 772 GVFLTQNVSDHLSSSAFMSLAARPPKPLSSREDERNVRSVVVEDPE--CCILNL-----N 825
DB 1207 -----KKLESS-----EENLSS-----EDEELPCFOHLLFGKVN 1235

QY 826 EIPSQEKVQHPSDMEVSGVDGSKSE-----QLRDCSNSGIERNFLEKSIQNLEBEV 878
DB 1236 NIPS--QSTRH--STVATECLSKNTENLLSLKNSLNDCSNQVI-----LAKASQ-----EHH 1284
QY 879 LSSQDSFDPALFQSCGRVSGSCSKSDAEFFPTTRCETKTVSGTSQSVQGTGSPN-----LSD 934
DB 1285 LSEETKCSASLFS-----QCSELEDLTANTNTQDPFLIGSSKOMRHOSESQGVGLSD 1337
QY 935 EICLOGNERPHLYEGSGDVQKQETTNVAKKPDLEKLTWNKDSVCFGQPRNDTNNQTTPS 994
DB 1338 KELVSDDEE-----RGTG-----LEENN--QEQSDMSNLGEAASGC-----ESET 1376
QY 995 SSYEOCATRQPHVLDIEDFGMQGEGLYGWSMISPRVDRVKNKNVPRFRFGGSGVPREF 1054
DB 1377 SVSEDCS-----GLSSQSDILTTQQRDTMQHN-----LIKLOQEMAELEAVLEOH 1421
QY 1055 TCQIIPSTPHELPGMGLSGSSSAVOEHODDT--OHNQODEMNKASHLOKTFDLDLLNSSEEC 1113
DB 1384 -----GLSSQSDILTTQQRDTMQHN-----LIKLOQEMAELEAVLEOH 1421
QY 1114 LTROSSTKQNTDGLCLPRDRTAEDVVDPLSNSSNQILVSNSSNKBQTAVEYKETNAT 1173
DB 1422 GSOPNSYPSI-----ISDSSALEDLRNPEQSTSEKAVLTSQKSSEYP 1464
QY 1174 ILREMGTLADGKPTSQWDSLRLKDVNEGREGQRERNKNMDSIDYEAIRRASISISEAI 1233
DB 1465 ISQNPGLSAD--KFEVSADSTSKNKE--PGVERSSPSKCPSLD-----DRWYMHSCSGSL 1517
QY 1234 KERGM--NNMLAVRIKDFLERIVKDHGGIDL 1262
DB 1518 QNRNYPQOEELIKVYDVEEQOLEESGPHDL 1547

Search completed: September 27, 2002, 12:55:48
Job time: 9577 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 10:26:26 ; Search time 70.37 Seconds
(without alignments)
2360.927 Million cell updates/sec

Title: US-09-840-743-2
Perfect score: 9089
Sequence: 1 MOSIMSSAVNATEAQND.....PRPLMARLHPASKLNKNT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3949	43.4	1017	2 T48452	hypothetical prote
2	2863	31.5	555	2 T48453	hypothetical prote
3	1559.5	17.2	1207	2 D84781	hypothetical prote
4	1260	13.9	234	2 T48454	hypothetical prote
5	709	7.8	917	2 T03430	hypothetical prote
6	278	3.1	5327	2 T13564	microtubule-associ
7	246.5	2.7	2897	2 B48666	cell proliferation
8	246.5	2.7	3256	2 A48666	cell proliferation
9	240.5	2.6	2938	2 T30249	hypothetical prote
10	233.5	2.6	2218	2 B4683	endonuclease III -
11	232	2.6	259	2 D75275	probable heat choc
12	219.5	2.4	1871	2 D96796	glucan 1,4-alpha-g
13	217.5	2.4	1490	2 T20513	hypothetical prote
14	217	2.4	2022	2 T48818	hypothetical prote
15	211	2.3	1188	2 T05324	hypothetical prote
16	210.5	2.3	1805	2 A34736	hypothetical prote
17	208.5	2.3	3924	2 S37431	nestin - rat
18	207	2.3	3331	2 T25410	ankyrin 2, neurona
19	206	2.3	3187	2 JC5837	hypothetical prote
20	205.5	2.3	2562	2 T14266	364K Golgi complex
21	205	2.3	1983	2 T00385	Xin protein - chic
22	204.5	2.2	2717	2 A34203	KIAA0624 protein -
23	197.5	2.2	1145	2 T18235	DNA-binding protei
24	197	2.2	1366	2 B86292	transcription acti
25	196.5	2.2	3225	2 I52300	F7H2.12 protein -
26	196	2.2	3488	2 T34418	giantin - human
27	192.5	2.1	2954	2 T14156	hypothetical prote
28	192.5	2.1	3259	1 A56539	kinesin-related pr
29	192	2.1	1888	2 T14273	giantin - human
					zinc finger protei

ALIGNMENTS

RESULT 1

T48452
hypothetical protein T32M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48452
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224487
A:Accession: T48452
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1017 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 167/1; 874/1
A:Note: T32M21.160

Query Match	43.4%	Score	3949;	DB	2;	Length	1017;
Best Local Similarity	99.9%	Pred. No.	1.1e-206;				
Matches	755;	Conservative	1;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MOSIMSSAVNATEAQNDGSRQDYLEFDLNTPOQKFSKRKRKMPKVVYEGKPKRP	60				
DB	235	MOSIMSSAVNATEAQNDGSRQDYLEFDLNTPOQKFSKRKRKMPKVVYEGKPKRP	294				
QY	61	RKPAELPKVVYEGKPKRKKAATQEKVSKETGSAKKNLKESATKKPANNYDMSNKP	120				
DB	295	RKPAELPKVVYEGKPKRKKAATQEKVSKETGSAKKNLKESATKKPANNYDMSNKP	354				
QY	121	EYTLSCRKALFNDELNPGDARQDSESEIVQNSSGANSFSEIRDAIGTNGSFLDSVSQ	180				
DB	355	EYTLSCRKALFNDELNPGDARQDSESEIVQNSSGANSFSEIRDAIGTNGSFLDSVSQ	414				
QY	181	IDKTNGLGAMNQLFVSMGNQPKLSTGAKLARDQDPDLTRNQOCQFPVATONTQFME	240				
DB	415	IDKTNGLGAMNQLFVSMGNQPKLSTGAKLARDQDPDLTRNQOCQFPVATONTQFME	474				
QY	241	NOQAWLQKLNQLIGFFGNOQPMIRNQOCPCLANGNQPMYLLIGTPRALYSGNQQLGG	300				
DB	475	NOQAWLQKLNQLIGFFGNOQPMIRNQOCPCLANGNQPMYLLIGTPRALYSGNQQLGG	534				
QY	301	POGKRPITFLNHQTCPLPAGNOLYGSPTDMHOLVMSGTGGQHQGLLTKNQOGLSLRGQGPC	360				
DB	535	POGKRPITFLNHQTCPLPAGNOLYGSPTDMHOLVMSGTGGQHQGLLTKNQOGLSLRGQGPC	594				
QY	361	VPLIDQOPATPKGFTHLNQWATSMSSPGLRPHSOSQVPTTYLHVESVRLNLTGTCQ	420				

Db 595 VPLIQOAPATPKFTHLNOMVATSMSSPELRPHSOSQVPTTYLHWESVRLINGTTGTQ 654
 QY 421 RSRAPAYDSLOODIHQGNKYILSHEISNGCGKALPONSILPTPIMLAEARSGKROY 480
 Db 655 RSRAPAYDSLOODIHQGNKYILSHEISNGCGKALPONSILPTPIMLAEARSGKROY 714
 QY 481 HRAMQTEKHDLNLAQOIAQSDOVERHNSSTCEVYLDAAKTKIKOVVOENJHGMPEVI 540
 Db 715 HRAMQTEKHDLNLAQOIAQSDOVERHNSSTCEVYLDAAKTKIKOVVOENJHGMPEVI 774
 QY 541 EIEDPTGARGKNTASISKSGKNSPVKTAKEKCTYPKPAKGRGKRSKSV 600
 Db 775 EIEDPTGARGKNTASISKSGKNSPVKTAKEKCTYPKPAKGRGKRSKSV 834
 QY 601 PAHASEIOLMOTPPKPTPLSRKSPKGRKSIOGSKARGPSGELLCDOSIAEIIYRMON 660
 Db 835 PAHASEIOLMOTPPKPTPLSRKSPKGRKSIOGSKARGPSGELLCDOSIAEIIYRMON 894
 QY 661 LYLGDKEKEQONAMVLYKGGALVPIESKRRKPRKYDIDETTRINMLMGKDEKEG 720
 Db 895 LYLGDKEKEQONAMVLYKGGALVPIESKRRKPRKYDIDETTRINMLMGKDEKEG 954
 QY 721 DEEKDKKKKKWEEERVRFRGRADSFIAHMLVQGD 756
 Db 955 DEEKDKKKKKWEEERVRFRGRADSFIAHMLVQGE 990

RESULT 2

T48453
 hypothetical protein T32M21.170 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_rev1sion 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T48453
 R:Beyan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De
 ewes, H.W.; Ridd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, March 2000
 A:Reference number: Z24487
 A:Accession: T48453
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-555 <BEV>
 A:Cross-references: EMBL:AL162875
 A:Experimental source: cultivar Columbia; BAC clone T32M21
 C:Genetics:
 A:Map position: 5
 A:Introns: 469/3; 496/2; 524/3
 A:Note: T32M21.170

Query Match 31.5%; Score 2863; DB 2; Length 555;
 Best Local Similarity 100.0%; Pred. No. 5e-148;
 Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SSAPMSIAARPPKLTSSREDEERNVRSVVDPECCILNLNLEIPSMOKVQHPMSMEVSG 844
 Db 7 SSAPMSIAARPPKLTSSREDEERNVRSVVDPECCILNLNLEIPSMOKVQHPMSMEVSG 66
 QY 845 VDSGSKEDLRDCNSGIERFNFLEKSIONLEEEVLSODSDFPAFOCGVSGVSCSCSKS 904
 Db 67 VDSGSKEDLRDCNSGIERFNFLEKSIONLEEEVLSODSDFPAFOCGVSGVSCSCSKS 126
 QY 905 DAEPPTTTCETKTVSGTSQVOTGSPNLSDICLOGNERPRLYEGSGDVQKQETTTNAOK 964
 Db 127 DAEPPTTTCETKTVSGTSQVOTGSPNLSDICLOGNERPRLYEGSGDVQKQETTTNAOK 186
 QY 965 KPDLKMTNMKDSVCFQPRNDTMMQTTTPSSSYECACATROPHVLDIEDFGKQGBGLYSW 1024
 Db 187 KPDLKMTNMKDSVCFQPRNDTMMQTTTPSSSYECACATROPHVLDIEDFGKQGBGLYSW 246
 QY 1025 MSISPRVDVAVKNNVPRRFQGGSVPEFTGOLIPSTPHLPGLMGSSGSSAVQEHODD 1084
 Db MSISPRVDVAVKNNVPRRFQGGSVPEFTGOLIPSTPHLPGLMGSSGSSAVQEHODD 306

QY 1085 TQHNQDEMNKASHQKTFLLDNLNSELCTROSSSTKONTIDGCLPRDTADVDVPLSN 1144
 Db 307 TQHNQDEMNKASHQKTFLLDNLNSELCTROSSSTKONTIDGCLPRDTADVDVPLSN 366
 QY 1145 NSSLONILVESNSSKKEOTAVYEKETNATILREMGTLADGKPPISQWDSLKKDYBNGE 1204
 Db 367 NSSLONILVESNSSKKEOTAVYEKETNATILREMGTLADGKPPISQWDSLKKDYBNGE 426
 QY 1205 ROERKNNMDSIDYPAIRASISEISEAIKERGMNMLAVRLKDLERIVKHGGIDLEW 1264
 Db 427 ROERKNNMDSIDYPAIRASISEISEAIKERGMNMLAVRLKDLERIVKHGGIDLEW 486
 QY 1265 LRESPDRAKDYLLSTIRGLGKSVCEVRLTLHNLAFPDVTNAGRIAVRMGWVPLQPLPE 1324
 Db 487 LRESPDRAKDYLLSTIRGLGKSVCEVRLTLHNLAFPDVTNAGRIAVRMGWVPLQPLPE 546
 QY 1325 SLQHLLEL 1333
 Db 547 SLQHLLEL 555

RESULT 3

D84781
 hypothetical protein At2g36490 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_rev1sion 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: D84781
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 euss, D.; Merzhan, M.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventler
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: D84781
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1207 <STO>
 A:Cross-references: GB:AE02093; MID:94581149; PIDN:AMD24633.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g36490
 A:Map position: 2

Query Match 17.2%; Score 1559.5; DB 2; Length 1207;
 Best Local Similarity 28.5%; Pred. No. 1.1e-76;
 Matches 480; Conservative 199; Mismatches 366; Indels 641; Gaps 61;

QY 13 TEATEQNDGSRQDVLFDLNTPOQKPSKRRKFRMPKVVVEGKFRKPRKPAELPKVVE 72
 Db 65 TEEVESLSVSNVAE-QILKTP-EKPRKKHR--PKVREAKPRKREKPRAPRKSVTVD 120
 QY 73 GKPRKPRKATQOKV-KSETGSAKKNLKESATKPRANGDMSSNKPETLSCRAL 131
 Db 121 GQSKTPRKRYVRKKEVSKD-----QDATPEVSSAAVETSTRPK--RLGRVL 167
 QY 132 NFEDLEN-----PGDAR-GDSESEIVONSSGANSFSEIRDAIGTNGSFQDSQIDPTN 185
 Db 168 DFEADENENQNGSIRREAGMESALQEKDLSGN-QELKCL-----LSAPTPRKRR 219
 QY 186 GLGAMNPLEVSMGNQDPK-----LSTGAKLARDQOPDLTLRN-----QOCQFP 229
 Db 220 SQGRK-----GVQPKKNGSNLEEVDISMAQAARKRQGPCTCDMLNSGIQYDEQCY- 271
 QY 230 VATONTQFPMENOAMLOMKNQILGFFPGNQPMTIRNOQPCCLAMGNQOPMYLIGTRP 289
 Db 272 -----QKMMWLYSPML-----QOGGM--RYDAIC----- 293
 QY 290 ALVAGNOOLGAPQGNKRPILFNHOTCLPAGNOLYGSPTDMHQLVMSYGGQOHGLIRNOQ 349
 Db 294 -----SKVFSQOAHNYSAF--HATCTSSISQLSANRY-----LTVERK 331
 QY 350 PGLIRGQQPCVPLIDQOAPATP-----KGFHLNOMVATSMSSPGLRPHSOSQV 398
 Db 350 PGLIRGQQPCVPLIDQOAPATP-----KGFHLNOMVATSMSSPGLRPHSOSQV 398

[illegible]

RESULT 4
T48454
hypothetical protein T32M21.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_c
C:Accession: T48454
A:Ref: Bevan, M.; Terry, N.; Ardiles, W.; Buyssshaert, C.; Dasso,
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224487
A:Accession: T48454
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <BE>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 49/2; 61/3; 81/1; 127/3; 149/3; 188/1; 201/1
A:Note: T32M21.180

Query Match 13.9%; Score 1260; DB 2; Length 234;

Best Local Similarity 100.0%; Pred. No. 1.9e-61;

Matches	234;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	----	--------	----	------	----

QY 1496 MTLREHMERNMELQEGDMSKALVALHPPTTSIPTPKLKNISRLRTEHQVVELPDSHRI.D 1555

1

DD I MTLREHMERNMELQEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVVELPDSHRLD 60

QY 1556 GMDKREDDPSYLLAIWTPGETANSAQPPEOKCGKASGKMCFEETCSECNSLREANSO 1615

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

DD 01 GMDKREPDDPSPYLLAIWTPGETANSAQPPEQKCGGKASGKMCDFDETCSECNSLREANSQ 120

QY 1616 TVRGTLIPCRTAMRGSFPLNGTYFQVNELFADHESLKPIDVPRDWIWDLPRRTVYFGT 1675

[illegible]

121 IVKSIJLLIFCRIMRGSTFLENGIIFQVNELFADHESLKPIDVPRDWIWDLPRTVYFGT 180

QY 1676 SVTSIFRGLSTEQIQFCFWKGFVCRGFEQKTRAPRPLMARLHFPASKLKNNKT 1729

db 181 SVTSTFRGLSTEOIOECEFHWKCEVCVRCEFAKTBADDDI MARI HEDACKY KNHNM 334

234

RESULTS

RESOUR
T05430

hypothetical protein F28A23.180 - *Arabidopsis thaliana*

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence revision 22-Apr-1999 #cont

Sequence_revision 23-Apr-1999 #lexc_change 24-Nov-1999
C;Accession: T05430

R; Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.;

submitted to the Protein Sequence Database, October 1998
A; Reference number: Z15415

A;Accession: T05430

A;Molecule type: DNA
A:Residues: 1-917 <REV>

THIS PAGE BLANK (USPTO)